UiO : Department of Informatics
University of Oslo

Evaluation of a Markerless Motion Capture System as a Tool for Sports Movement Analysis

Implications for ACL injury risk assessment

Tina Stenerud Moen

Master's Thesis Fall 2014

Evaluation of a Markerless Motion Capture System as a Tool for Sports Movement Analysis

Tina Stenerud Moen

November 3, 2014

Abstract

Background:

Marker based motion capture (MBS) is the preferred method for 3D motion analysis in sports injury prevention. There is a continuous search for more efficient and low-cost methods that can produce equally reliable measures. Markerless motion capture systems (MLS), like the Kinect v1 and v2 are cheap commercial 3D sensors that have implications for injury risk research. The purpose of this thesis is to validate a markerless 3D motion capture system (MLS) against a marker-based motion capture system (MBS).

Objectives:

- To evaluate the accuracy and tracking ability of MLS, Kinect v1 and v2
- To investigate the side tracking ability of MLS
- To investigate the inter-trial reliability of MLS for each task
- To assess methods for synchronizing the data from the MBS and MLS in time and space
- To consider use of MLS in a clinical setting for knee motion assessment

Material and Methods:

During two screening sessions with 12 (3) Norwegian elite female basketball players performed three tasks; a vertical drop jump, two-legged squat (front/side) and single-leg squat (left/right). The MLS (Kinect v1/v2) and MBS (Qualisys) simultaneously recorded the respective testing sessions in a 3D motion analysis lab. Kinematic measures were calculated and compared in a statistical analysis.

Results:

Kinect v1 and 2 had very good correlation scores for mean PF during the two-legged squat, indicating a significant relationship between MLS and MBS. Almost all VDJ variables were not significantly different from the MBS (two-tailed paired samples t-test, $p < 0.05$). KASR was poorly correlated for Kinect v1, except for mean PF during the squat(SRCC: 0.93) The two-legged squat scored highest overall in inter-trial reliability for all variables in session 1 and 2 (ICC: 0.81 - 0.99). FPPA was estimated more accurately by Kinect v1 than v2.

Conclusion:

Visual inspection of curves (PF and FPPA) for VDJ and squat, confirmed a real relationship between MLS and MBS variables. Adjustments to screening protocol will most likely improve the accuracy of the skeletal tracking (front and side) for Kinect v2 MLS. A simple method for estimating a common start-point in time for MLS and MBS, was found in session 1. Synchronization in space for kinematic measures require defining local coordinate systems for each body segment in MLS skeleton. Implementation of MLS in a clinical setting require further testing and verification of its accuracy in motion tracking.

Acknowledgments

Working on this master project was more fun than I could ever imagine, but that is only thanks to everybody that has helped me and showed interest in my work. First of all I have to thank my supervisors Asbjørn Berge at SINTEF and Tron Krosshaug at NIH, for giving me the opportunity to apply my technical experience to the field of sports sciences that was already a passion of mine.

My biggest thank you goes to Kam Ming Mok at OSTRC for teaching me how to formulate and plan my research, and for always being there when I needed help. I want to thank all the basketball girls for volunteering to be a part of my experiments, and Dita for her help with last minute test preparations. Thank you Oliver for bringing the Kinect sensor from Germany. Thank you Jeanine, my good friend and physiotherapist who was the best research assistant I could ask for.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This master thesis is done in cooperation with SINTEF IKT and researchers at the Oslo Sports Trauma Research Center (OSTRC). The work presented is motivated by the implications that recent development of powerful three-dimensional (3D) sensors have on analysis of human sports movement and injury prevention research. Motion Capture is the name of the field that covers the many techniques for creating a digital model of a moving object or person in a scene. There are several areas of application for the motion capture technology; the most obvious is the movie and entertainment industry. Avatar is one example of a film that uses motion capture technology to create computer animations of people and objects.

1.1 Motion Capture in Health and Sports Applications

Sports performance is a field of application where motion capture tools for evaluating and improving performance are advantageous to coaches and general managers, as well as players. They all have an interest in analysis that goes beyond looking at two-dimensional video recordings. Scientists use motion capture systems and techniques in a wide range of research fields, including medical and clinical analysis. Reflexion Health[1] out of San Diego is a software company that makes use of the Kinect sensor in digital applications for physical rehabilitation programs. Motion analysis using motion capture systems are common in sports training and assessment of human kinematics, and an important tool in rehabilitation. These systems require a laboratory setup with multiple high-frequency cameras. The cameras use infrared technology and rely on retro-reflective markers placed on anatomical landmarks on the person being recorded. The location of the markers can be found in three dimensions by the cameras using triangulation. Dedicated software estimates the human skeleton with a high degree of accuracy, but are prone to be affected by marker displacement caused by skin or muscle artifacts. Vicon and Qualisys are examples of companies delivering hardware and software for current marker based motion capture systems. Implementing marker based motion capture in motion analysis projects is an ex-

[1] See http://reflexionhealth.com/

pensive and time-consuming process, and there is a high demand for markerless systems that are easier to implement in clinical settings [27].

The last few years there has been an increase in availability of motion capture devices in consumer electronics used in video gaming. This is made possible by the constant development of innovative and less expensive ways of capturing and translating human motion into a realistic 3D model. These markerless motion capture systems (MLS), like the Microsoft Kinect sensor for Xbox (PrimeSense Technology), have implications for other application areas such as healthcare and medical research. Kinematic measurements obtained from a MBS are subjected to inconsistency in placement of the markers on a session to session basis. According to the review by McGinley in 2009 [24] this was the main source of error, when measuring the variability in repeated kinematic measurements in that study. A markerless approach can eliminate the error caused by inconsistent marker placement, if the exact same model is applied for each repeated session, and without any change in the shape of the body of the person being tested. A MLS uses only one model which can be a source of error due to the numerous types and sizes of human bodies that exist. In order to deal with this type of error, the Microsoft Kinect has used a very large set of synthetic and real training data in their skeleton tracking algorithm. The algorithm uses a statistical approach to estimate the joint positions, and this method which will be presented in more detail later.

1.2 Human Pose Recognition

In computer vision the process of recognizing human poses in images is called articulated pose estimation. For a sequence of images this becomes more challenging and is known as articulated tracking. The main objective is to see how the body pose changes over time, for example through a series of video image frames. Tracking and estimating human body poses in scenes and settings, that do not require restrictions on lighting human appearance, light and complexity of pose, is a challenge. It is difficult to reconstruct 3D objects and scenes using only a single camera. Obstacles include occlusion of body parts and ambiguities related to depth and reflection. To get improved accuracy in body estimation, additional cameras or other sensing techniques are needed. The previously mentioned Xbox Kinect sensor is an example of motion capture device with a sensor technology that finds 3D configurations by emitting a pattern of infrared structured light. This makes it easier to estimate the depth of the objects in the scene. Shotton et. al [36] presented two methods using single depth images (such as the ones that can be obtained from the Kinect) to estimate human poses. Although the Kinect depth sensor overcomes some of the challenges related to human pose estimation, it also has it limitations. The infrared structured light can be distorted by sunlight and is therefore not applicable for use outside. The skeletal tracking delivered with the Kinect software, experiences difficulties with sideways poses where for example one leg occludes the other [37].

The Kinect sensor is potentially a good markerless alternative for motion analysis in sports trauma research. Assessment of joint kinematics and kinetics of the lower extremities for athletes, the most accurate and also most expensive

tool is using 3D laboratory-based motion analysis [30]. In 3D motion-analysis for the purpose of injury-risk analysis a large number of reflective markers have to be attached to the legs, arms and torso of the test-subject. The athlete then performs various sports-related movements as realistic as possible, while the marker trajectories are recorded by numerous infrared cameras. The cameras capture the reflection of infrared light. The position of each marker is calculated with high spatial resolution by the time-of-flight information received. This is an extremely time-consuming and expensive process that is limited to one location and cannot easily be transferred to the consumer-market. Only some research centers have the funds to support building a lab with the space and equipment necessary for marker based motion analysis.

In 2014 Microsoft released the Kinect v2 sensor as a part of the upgraded gaming system called Xbox One. The Kinect v2 is based on parts of the same time-of-flight principle as the MBS, to estimate depth and reconstruct 3D objects. The upgraded sensor has a larger field of view which allows the player to stand closer to the sensor and still achieve full body tracking. A closer look on how the new version of the sensor is able to more effectively capture depth information to track human pose, follows in the next chapter.

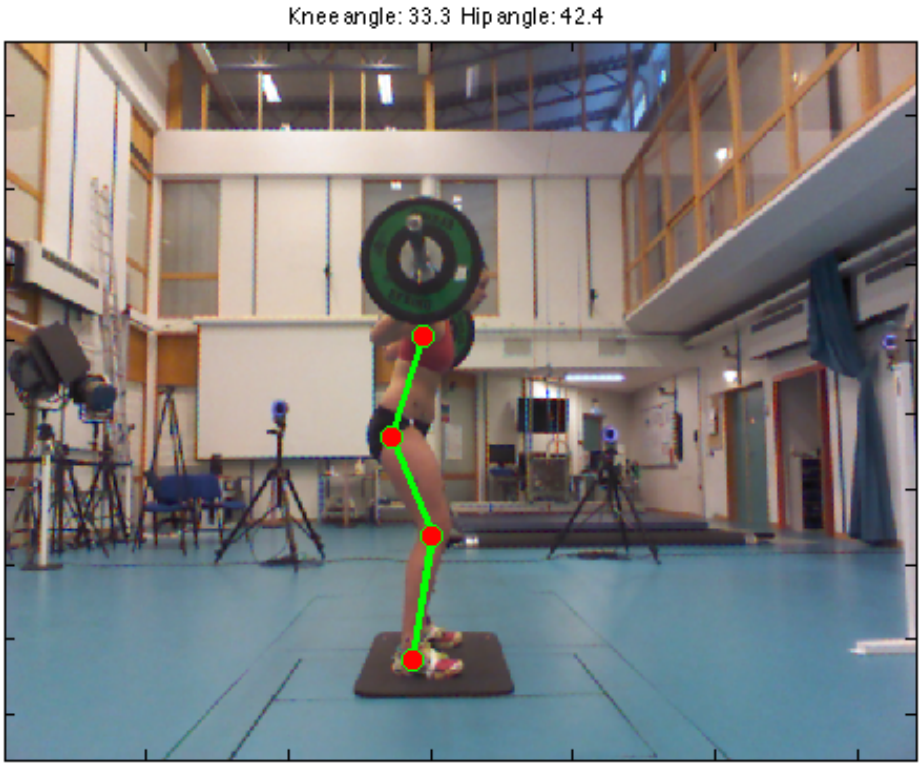

Figure 1.1: From the 3D motion analysis lab at OSTRC. The red circles indicates joint positions found with a tracking algorithm, and the green lines represent the bone segments. The tracking algorithm will be explained in chapter 3.

1.3 3D Motion Analysis in Sports Injury Research and Rehabilitation

The Oslo Sports Trauma Research Center (OSTRC) at the Norwegian School of Sport Sciences, does research on several areas related to sports injury prevention. The research includes finding risk factors and injury mechanisms for serious and common injuries, such as tearing of the ACL and other knee injuries. Sports injury prevention is a field that benefits from the use of motion capture in its research. A image from the lab at OSTRC during screening session 1 is seen in figure 3.11 on page 57. The red dots indicates joint positions found with tracking algorithm, and the green lines represent the bone segments. The tracking algorithm will be explained in chapter 3. At OSTRC there is an ongoing study which investigates whether anatomical, biomechanical and neuromuscular features can be predictors of future ACL injury. The projects related to screening athletes for ACL injury risk at OSTRC, rely on the marker based motion capture system installed in a biomechanics lab. The MBS is well-known as the superior method in regards to validity and reliability. There is a continuous search for more efficient and low-cost methods that can produce equally reliable measures in this field of application [25]. Easier access to a simple screening method that is not too time-consuming or expensive, implies that a larger population of athletes can be tested for ACL injury risk. Screening programs for the purpose of injury prevention would benefit vastly from using a MLS, and could possibly be implemented in a smaller clinical setting.

It is becoming more and more common in research to make use of MLS, like the Kinect sensor, to quantify movements within sports, rehabilitation and injury prevention. Because few studies have tested the reliability and validity of these markerless motion capture systems, they are not extensively used as a screening tool for ACL injury. Marker based motion capture (MBS) is the preferred method for validating markerless methods. The MBS methods are used to analyze motion that requires a larger field of view than methods based on radiography and bone pins. A recent study by Pfister et al. [34] on the comparative abilities of Microsoft Kinect and Vicon 3D motion capture in gait analysis, concluded that Kinect has basic motion capture abilities but large variability in measurements for higher walking/jogging velocities. When the markerless technology becomes as accurate as the marker based motion capture, it will be necessary to validate the MLS against radiographic methods for better accuracy estimation.

1.4 Objectives

- To evaluate the accuracy and tracking ability of MLS; Kinect v1 and v2
- To investigate the side tracking ability of MLS
- To investigate the inter-trial reliability of MLS for each task
- To assess methods for synchronizing the data from the MBS and MLS in time and space
- To consider use of MLS in a clinical setting for knee motion assessment

The first main objective is related to our application of choice; screening female athletes for ACL injury risk. We aimed to assess the accuracy of MLS by quantifying the difference in 3D kinematic measurement between the MLS (Kinect v1 and v2) and the MBS (Qualisys). More specifically this thesis evaluated MLS (Kinect) a tool for analyzing the following;

1. Knee motion during vertical drop jump (VDJ)
2. Knee motion during front-oriented squats
3. Knee motion during single-leg squat (SLS)
4. Knee and hip motion in the sagittal plane during side-oriented squats

Chapter 2

Theory

This chapter gives a more detailed overview of the technology used in modern 3D motion capture sensors. An explanation of theory and previous work that has been the motivation for this thesis, follows the technical descriptions. Although this chapter is written with an audience with similar academic background as myself in mind, it is meant to give any reader of this interdisciplinary master project a good overall understanding of both sides. Using a commercial markerless motion capture system (MLS) in injury risk research applications is very exciting; and the theory of biomechanics is closely related to principles used in the field of computer vision and robotics. This combination of MLS and biomechanics is almost revolutionary in the sense that for so many years there has been expressed a need for a reliable, easy-to-implement and cost-saving system that can live up to the standard of marker based motion capture systems. It remains to see if the technology is good enough to replace what is established as the ground truth, and this thesis will in part reveal how close the most up-to-date technology is to achieving this.

The focus in the following sections is on the specifications and features of the two versions of 3D imaging sensors, that Microsoft has released for the Xbox gaming systems under the name of Kinect. The Kinect is essentially a 3D motion capture device with a camera and a depth sensor that makes it capable of creating a model of the human body in three dimensions using infrared laser light. Since the Microsoft Kinect is a commercial product, a complete technical description of the sensor has not been publicly released. Except for investigations of the hardware components there is no official documentation that give full disclosure of the technology used in the system. Patents by Freedman et al. titled "Distance-varying illumination and imaging techniques for depth mapping" [16] and "Depth mapping using projected patterns" [17], were published in 2010 and describe methods for creating a three-dimensional depth map of objects in a scene. These patents partly give information about the technical specifications of the Kinect v1 sensor. The recent release of the Kinect v2 sensor that accompanies the new upgraded Xbox One gaming system, was of particular interest for this thesis as it was rumored to have improved features, including a more stable skeletal tracking. At this time there is not much documentation or literature available on the performance of the Kinect v2 sensor, but there are several known upgraded specifications compared to the first version of the

sensor. In the rest of this thesis I will specify *v1* or *v2* of the Kinect sensor unless the version is irrelevant for the topic explained.

2.1 Creating a Digital 3D Model

Computer vision involves use of any kind of imaging system that intends to transfer knowledge about the physical scene back to the computer using images, and this is where motion capture becomes useful. The intention is to reconstruct a digital three-dimensional world, but doing this using traditional two-dimensional cameras becomes challenging when you have problems related to perspective distortion and depth perception. It is difficult to determine the distance to an object with a single camera or image in a scene without knowing the actual size of the object. One solution to this problem of finding the depth of objects in a scene, is stereo-imaging with two cameras that capture the scene from two different angles. The depth is then calculated from triangulation the same way the human eyes work together to perceive depth in the real world. Objects further away have a smaller disparity in position than objects that are closer when observed from two different viewpoints. This displacement is called parallax and objects that are close have a larger parallax and appear to move faster than objects that are further away when viewed from a pair of stereo cameras [41]. The definition of parallax in the stereo-imaging technique is illustrated in figure 2.1.

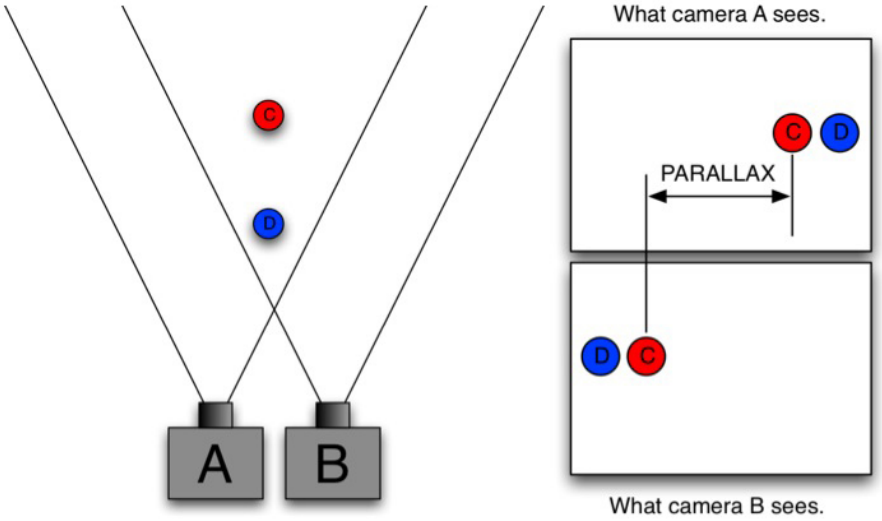

Figure 2.1: Stereo-imaging with two cameras and the different viewpoints that illustrates the concept of parallax [23].

2.1.1 Kinect - A Markerless Motion Capture System

The Microsoft's Kinect sensor makes use of triangulation and structured light to reconstruct a 3D-scene. The hardware components of the first Kinect sensor are illustrated in figure 2.2 on the next page.

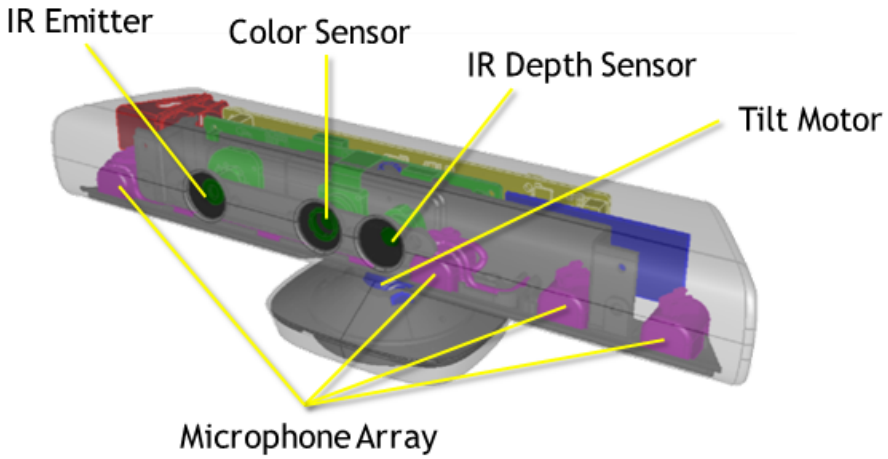

Figure 2.2: The Microsoft Kinect v1 components; IR emitter, IR depth sensor, color sensor (RGB camera), tilt motor for changing viewing angle and four-microphone array [31].

Structured Light Principle

The infrared(IR) emitter creates a speckle pattern by projecting laser light that is reflected back to the IR depth sensor. This technology is based on what is known as the structured light principle, which is a process where a known pattern of pixels is projected onto a scene. The pattern can be a grid or horizontal lines, but the infrared laser on the Kinect sensor projects a partly randomized pattern of dots.

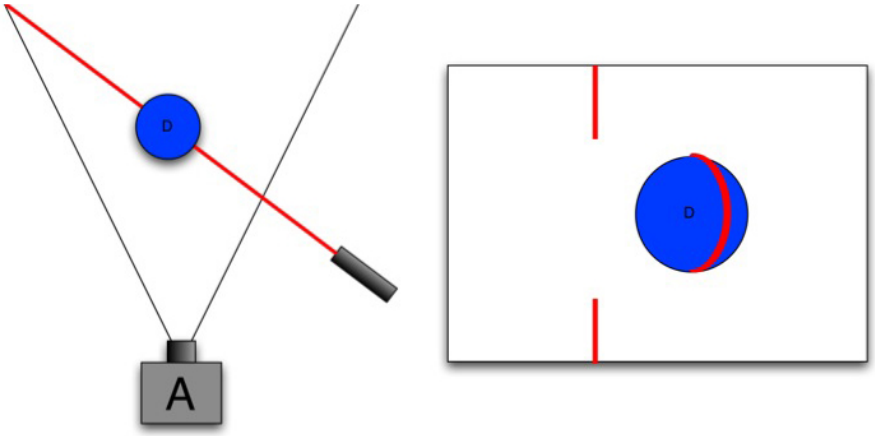

Figure 2.3: Structured-light imaging method illustrated with a single laser stripe. Objects that are close to the camera will have laser stripes that are more to the right than objects further away [23].

The structured-light principle is illustrated in figure 2.3 for one frame with one laser light stripe. Objects that are close to the camera will have laser stripes

that are more to the right than objects further away. This is explained in an online article by 3D imaging expert Daniel Lau. He writes that due to expansion of the projected light with distance from the sensor, the depth of the objects can be determined by looking at how far the dots have moved from left to right when comparing the incoming and outgoing pattern [23].

A single laser stripe is not sufficient to produce multiple points for each row in a camera with resolution 640x480. By sweeping the laser more points are generated and in order to account for objects that are moving, the Kinect sensor projects a pseudo-random infrared pattern that illuminates the whole scene. The dots are captured by the sensor and processed in order to match it with the transmitted pattern. The figure in 2.5 on the next page is taken from an article written by Zhang [44] at Microsoft Research and shows the IR dots seen by the IR sensor with a close up of the area in the red box. A depth map can then be constructed by matching the pattern from the IR emitter with the pattern that is observed by the IR camera. The reflected light is deformed when it hits the objects in the scene and with knowledge about the properties of the emitted pattern, it is possible to calculate the distance of the object from the sensor based on the deformation of the laser light [44]. The three-dimensional coordinates for the surface of the object is contained in the depth map. The depth image captured by the Kinect sensor in figure 2.4 is of the same scene as the image in figure 2.5 on the next page.

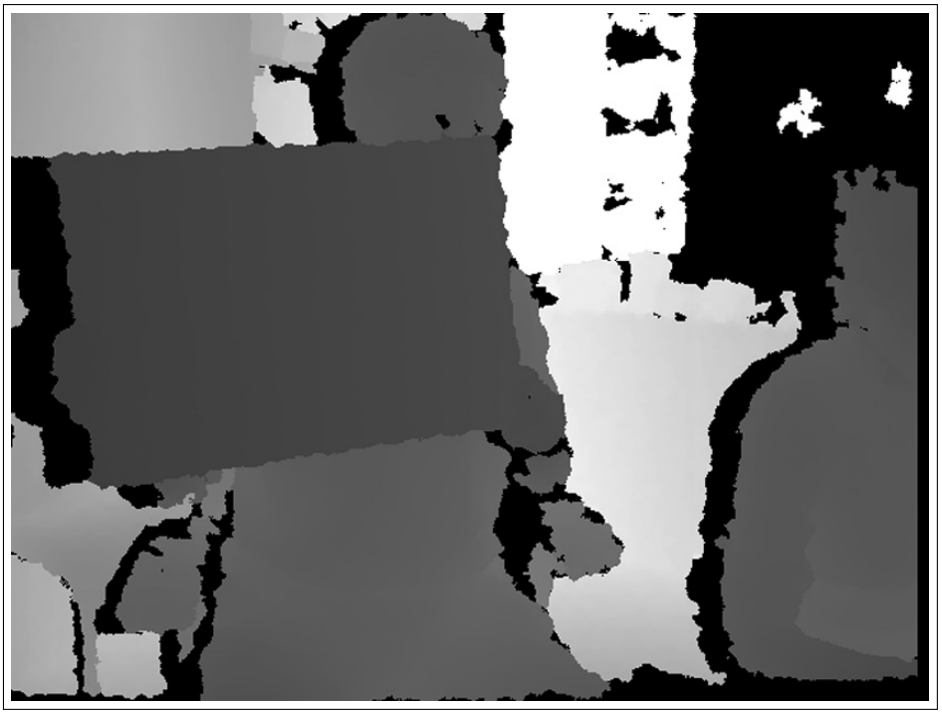

Figure 2.4: Kinect sensor depth image corresponding to the IR image in figure 2.5 on the next page [44].

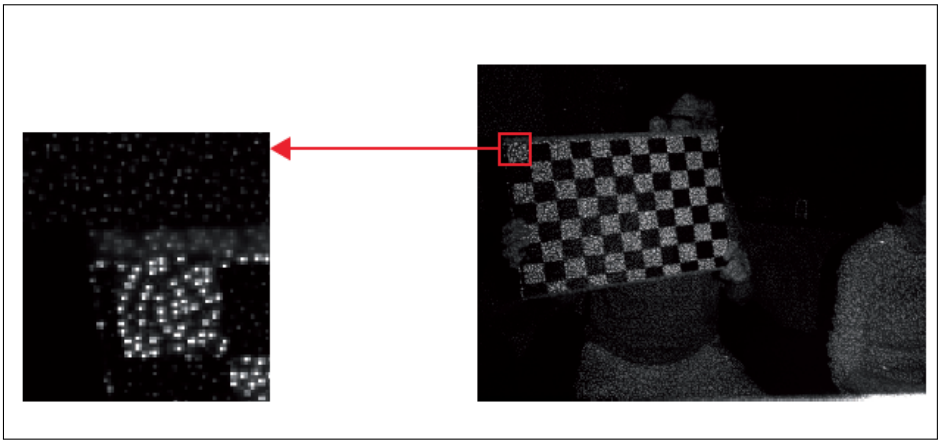

Figure 2.5: IR dots in the projected pattern seen by the IR sensor. The image on the left is a close up of the area in the red box in the image on the right [44].

The Kinect sensor uses a lens that is "astigmatic" which makes a projected circle (like in the speckle pattern) look like an ellipse because the focal lengths are different in the x- and y -directions. The orientation and direction of elongation of the ellipse are dependent on the depth of the object it is projected onto, so it varies with the distance from the sensor [16]. It is not sufficient to find one of the dots in the speckle pattern, but also its neighboring dots in a small window to correctly match it with the outgoing pattern. This causes problems for objects that are too small in the horizontal direction since the neighborhood of dots do not fit in the pixels of the object, and therefore depth cannot be estimated [23].

2.1.2 Microsoft Kinect v2

The biggest limitation of the first Kinect sensor is its inability to estimate depth for objects that only cover one image pixel in the captured image. In his post Lau also explains how the new version of the Kinect known as Kinect v2, uses a different method called time-of-flight that aims to solve this problem [23]. The imaging system in the Kinect v2 sensor measures distance based on the speed of light and the time it takes for a laser light pulse to travel from the IR emitter to the object and back to the IR sensor. The accuracy of the measurement of the time for the signal emitted from the sensor to make a round-trip, has to match the speed of light. Light Detection and Ranging (LIDAR) is the name of remote sensing systems used in examination of the Earth's surface that use pulses of laser light to measure distance and range with precision and accuracy. The same principle is used in the Kinect v2 sensor that has a near-infrared laser light source and camera with an indirect time-of-flight method. Estimations of the shape and depth of object in a scene come from the sensor's ability to capture the reflection of laser light off the objects. The Kinect v2 sensor can sense smaller objects than its predecessor by turning on and off half of a pixel so fast that it absorbs laser light only at the time it is turned on. Dividing the pixel in half makes it possible to delay the frequent turning on and off for the

second half with a phase difference of 180 degrees, which means that one half is always on while the other half is off. The infrared laser light is emitted as a pulse in phase with the first pixel-half meaning they will be "on" at the same time [23].

Indirect Time-of-Flight Method

Figure 2.6 illustrates the method the Kinect v2 sensor uses to indirectly measure the time-of-flight for the laser light pulse. The red areas indicate the light pulse and the grey columns are the time periods for when the first half of the camera pixels are turned on. In the first row the laser light pulse arrives at the same time as the first pixel-half is turned on and therefore it can absorb the light. At the same time the second half of the pixel will reject the light. The second row shows the offset when the laser light is at a small distance from the receiving camera pixels and the first half of the pixel will only absorb parts of the photons in the light before that half turns off. This time the second pixel-half will also absorb some of the photons in the laser when it turns off at the moment the first pixel-half turns off.

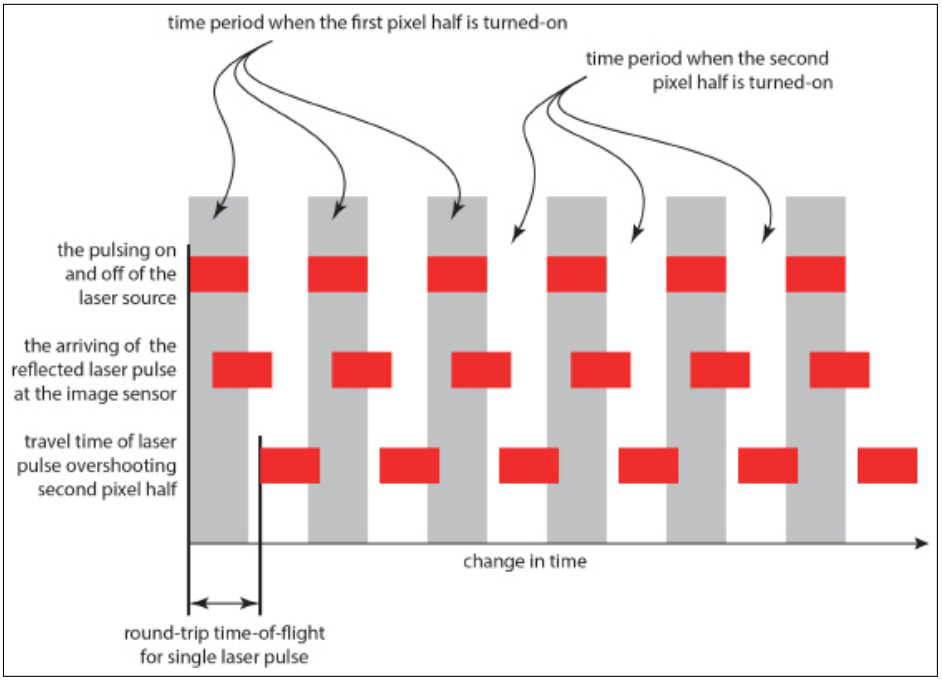

Figure 2.6: Indirect time-of-flight method used by the Kinect v2 sensor [23]

As the distance between the laser light emitter and the camera increases, more of the light photons will arrive at the time the second half is turned on. If more of the total amount of light after a short time period is registered by the second pixel-halves, it is assumed that the photons have traveled for a longer time which implies that the object is further away. To account for loss of energy

due to absorption of the light by the object that is hit, the ratio of the photons absorbed by the pixel-halves is registered and compared. Both pixel halves will be affected equally by the energy loss (fewer returning photons initially emitted), and therefore the ratio is a better measure for comparison than the number of photons. If the laser light travels a very long distance the photons can overshoot the second pixel-halves and arrive at the time the first halves have turned on again. The third row of red boxes in figure 2.6 on the previous page shows this case of ambiguity which can be eliminated by having the pixel-halves turned on for a longer time period. The disadvantage of this is that small travel distances will be harder to detect. The Kinect v2 sensor solves these problems as well by measuring the time-of-flight(TOF) twice to get two estimates of the distance. The first estimate is measured with an increased "on-time" for the camera-pixels that makes sure there is no ambiguity in distance. This estimate is then used to find and remove potential ambiguities in the second measurement which is a high-precision estimate. The degree of precision in the measurements is only limited by the speed of the sensor.

The Kinect v2 sensor is superior to the previous version in that it during a frame can automatically reject ambient light and reset the pixel. Ambient light sources like sunlight made it impossible to use the Kinect v1 sensor outside. Kinect v2 on the other hand can reject unwanted light which opens up for many more areas of applications.

The Kinect v1 sensor is as mentioned a camera sensor similar to other common stereo-imaging systems, but the newer time-of-flight method incorporated in the Kinect v2 sensor would normally be a lot more expensive if one were to buy it off-the-shelf. The advanced TOF technology comes in a very cheap package with the Kinect v2. Because it comes as a part of a gaming system, the Microsoft Cooperation can afford (and profit from) the manufacturing of such an advanced sensor.

The depth estimation in Kinect v1 has a depth resolution of maximum 1 cm since it as mentioned earlier depends using a window with neighboring pixels when evaluating the returned dot pattern. The time of flight for the laser light pulse in Kinect v2 is measured indirectly, and the distance for each pixel can then be determined individually. In the following section more details on the depth estimation method in relation to the body tracking ability of the Kinect sensor will be given.

2.1.3 Kinect Software Development Kit (SDK)

Microsoft has provided a Software Development Kit (SDK) to accompany a Kinect for Windows version of the sensor so developers can create applications or games that uses the Kinect's camera and microphone. There are only two versions of the Kinect sensor, which I have referred to as *v1* and *v2*, but several versions of the SDK which is updated on a regular basis.

The different versions of the SDK will be numbered; for Kinect v1 we have Kinect SDK 1.5 - 1.8, and for Kinect v2 we so far have Kinect SDK 2.0. The Kinect v1 sensor can identify as many as six people within in the field of view, but only track the detailed skeleton of two of them. A Kinect v1 application

has the ability to track users and their joints over time to register skeleton movement, but this is limited to poses where the user is facing the sensor. When the user is standing sideways, there are parts of the body that are not visible to the sensor and the skeleton will not be correctly recognized and cannot be tracked.

The IR sensor in the Kinect has a limited field of view which can make it impossible to obtain depth values for some pixels. If objects in the scene are too close or too far away, the pixels of those objects will appear black in the depth image since there are no computed values for those points. The vertical and horizontal field of view for the Kinect v1, in default depth range, is illustrated in figures 2.8 on the following page and 2.7. The Kinect v2 sensor has a larger field of view that makes it possible to stand closer to the sensor and still achieve full body tracking. The new version, Kinect SDK 2.0 for Kinect v2, also makes it possible to stream the infrared and the RGB images simultaneously, which was not possible in the previous version. To make sure that the players are able to use their arms, the recommended range is 1.2 - 3.5 meters from the sensor [31].

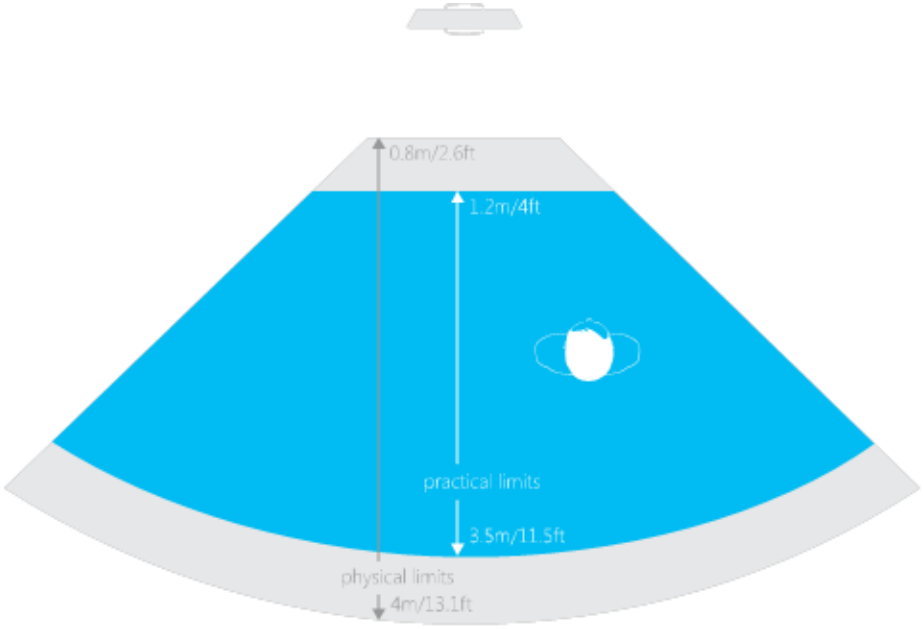

Figure 2.7: Horizontal Field of View for Kinect v1 showing the practical range and the sensor's physical limits [31].

Kinect Studio

Kinect Studio was released as a part of the Kinect SDK and comes in two different versions for each of the sensors. The Kinect v2 sensor and Kinect SDK 2.0 require Windows 8 or 8.1 as operating system in order to run applications. Screenshots of the two different versions of Kinect Studio are shown in figure 2.9 on the following page and figure 2.10 on page 28.

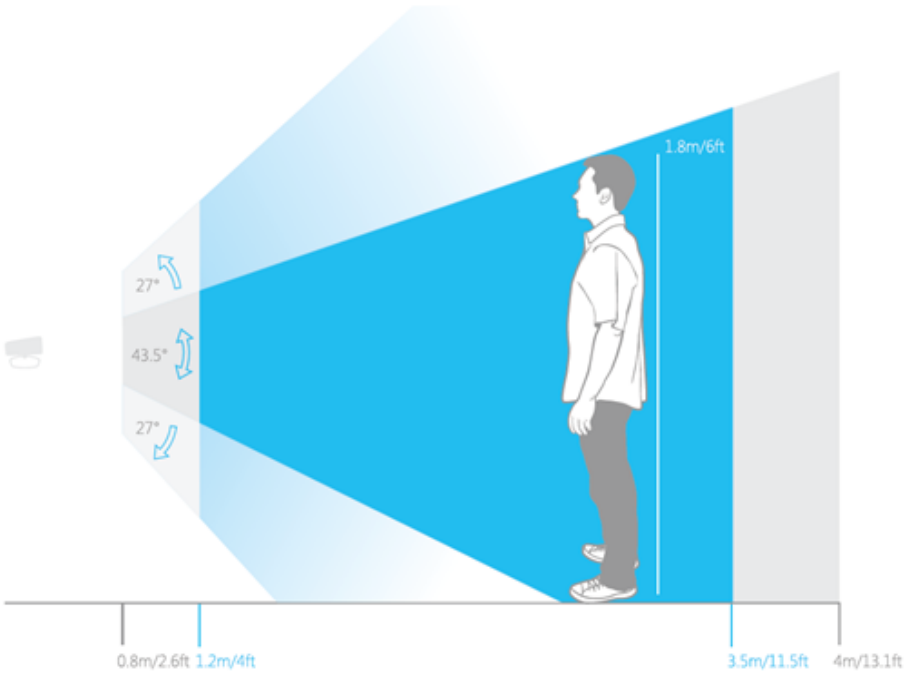

Figure 2.8: Vertical Field of View for Kinect v1 with an illustration of degrees of vertical tilting possible [31].

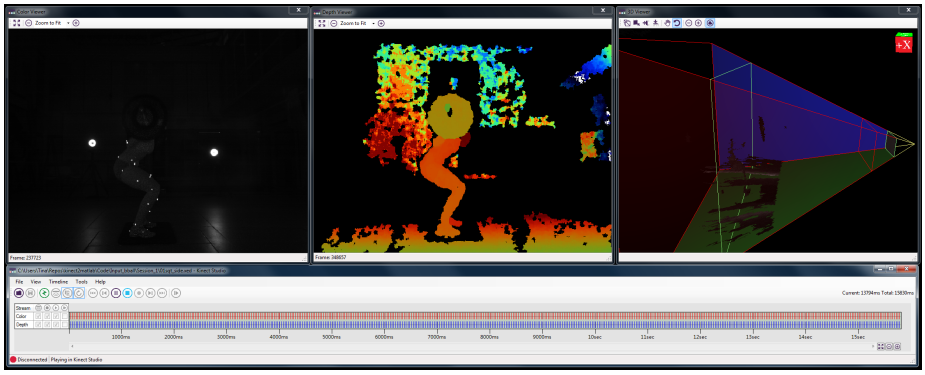

Figure 2.9: Screenshot of Kinect Studio 1.8 with the infrared view (left), depth view (center) and 3D view (right).

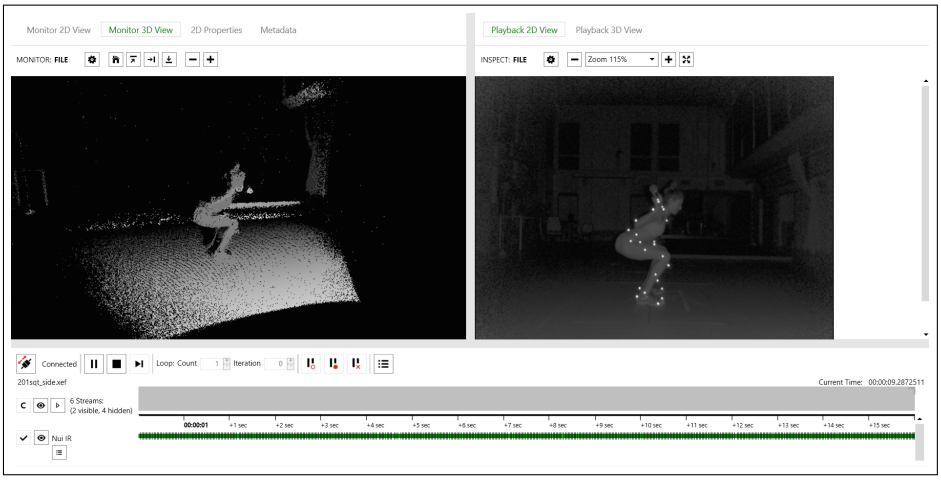

Figure 2.10: Screenshot of Kinect Studio 2.0 with the grey point cloud as the 3D view (left) and the 2D infrared view (right). The white spots in the infrared image are the retroreflective markers placed on anatomical landmarks for the marker based 3D motion analysis.

Open-source Alternatives

There are open-source libraries that enables the Kinect sensor to be used on other operating systems such as the OpenKinect project. The open-source driver (libfreekinect) for Kinect v1 and v2 is based on reverse-engineering and therefore does not have all access to all the features of the official SDK. So even if it is not optimal to use the Microsoft library, where you are stuck with Windows and Kinect Studio files that cannot be played back in other media applications, the official SDK is at least well-documented. Another advantage of sticking to the Kinect SDK, are the out-of-the-box applications that simplifies the process of getting started with developing for the Kinect sensor.

The example applications that accompany the SDK demonstrate the different uses and capabilities of the sensor. We adjusted one of these applications to fit our need in order to record raw data from the screening sessions for later processing in a Matlab program.

On a side note: Since the Kinect v2 sensor requires Windows 8 or 8.1, the open source driver recently made available for this sensor is much welcomed by developers that prefer other operating systems.

2.2 Tracking Human Motion

In the entertainment industry the problem of tracking human poses in real-time has been attempted solved in a variety of ways to give consumers an even more realistic and interactive gaming experience. The Nintendo Wii and Playstation Move entered the market with systems with hand held controllers that allowed the user to interact with the game using specific movements. Playstation Move uses triangulation with active markers in form of visible light in colored balls attached to the controllers. Nintendo Wii also relies on triangulation but with

infrared light. Microsofts Xbox 360 with Kinect was the first system that did not rely on controllers in order to have interactive gaming situations. When you have a real-time gaming system like the Xbox 360 it is a challenge to find a fast enough tracking algorithm that can track human pose regardless of the persons shape and size. Another problem is multi-player situations where several users have to be instantly recognized and tracked simultaneously. Two new methods for estimating human pose were presented by Shotton et al. [36] in 2011, and both methods work in real time and can track multiple bodies of different shapes and size. An overview of how the body joint positions are estimated in each of the methods is given in figure 2.11 on the following page.

2.2.1 Body Part Classification (BPC)

The first approach is the body part classification (BPC) algorithm that first extracts and labels the different body parts and then estimates the skeletal joint positions in 3D. A single depth image is used to segment out the human body and label each pixel to a body part based on classification using a deep randomized decision forest. Since each pixel is evaluated separately and labeled to a body part in each frame, the cost of searching all the joints is eliminated as the label for each body part specifies which skeletal joint it is localized in close proximity to. The body parts are recognized pixel per pixel and classified in a smaller window in the image. The 3D body joint positions of the skeleton is predicted by re-projecting the body parts to world space [37].

2.2.2 Offset Joint Regression (OJR)

The other method called "Offset Joint Regression" (OJR) is similar to the body part classification method, but instead of labeling the body parts at each pixel before the 3D joint position is predicted, it finds the location of the joints directly. The OJR approach uses a regression forest instead of the standard random forest that the BPC algorithm uses. The difference is that a regression forest has leaves that map the input variables to continuous outputs or in this case predictions for the joint positions. That means that from each pixel we get a "set of 3D offset votes" to the body joints without going through the intermediate step of labeling the body parts like in the BPC algorithm. The votes are used to create a density map for the world space and a local mean shift algorithm is implemented to find the modes of the density functions and the final set with proposed 3D joint positions [36]. Since the output is a set of predicted body joint positions in 3D, the algorithm gives an accurate estimation of the joint positions in one image frame and can also be used to track the positions in a sequence of images. This way the body tracking algorithm can initialize itself automatically in each frame and is not dependent on temporal information. Figure 2.12 on page 31 shows an example of how the OJR algorithm works when we have joints that are occluded and not in the field of view of the sensor. The pixels are illustrated as black squares and each of them has orange lines representing the 3D votes for the joint position. The set of joint position predictions are aggregated using the mean shift algorithm mentioned earlier, and the prediction with the highest confidence is seen in figure 2.12 on page 31.

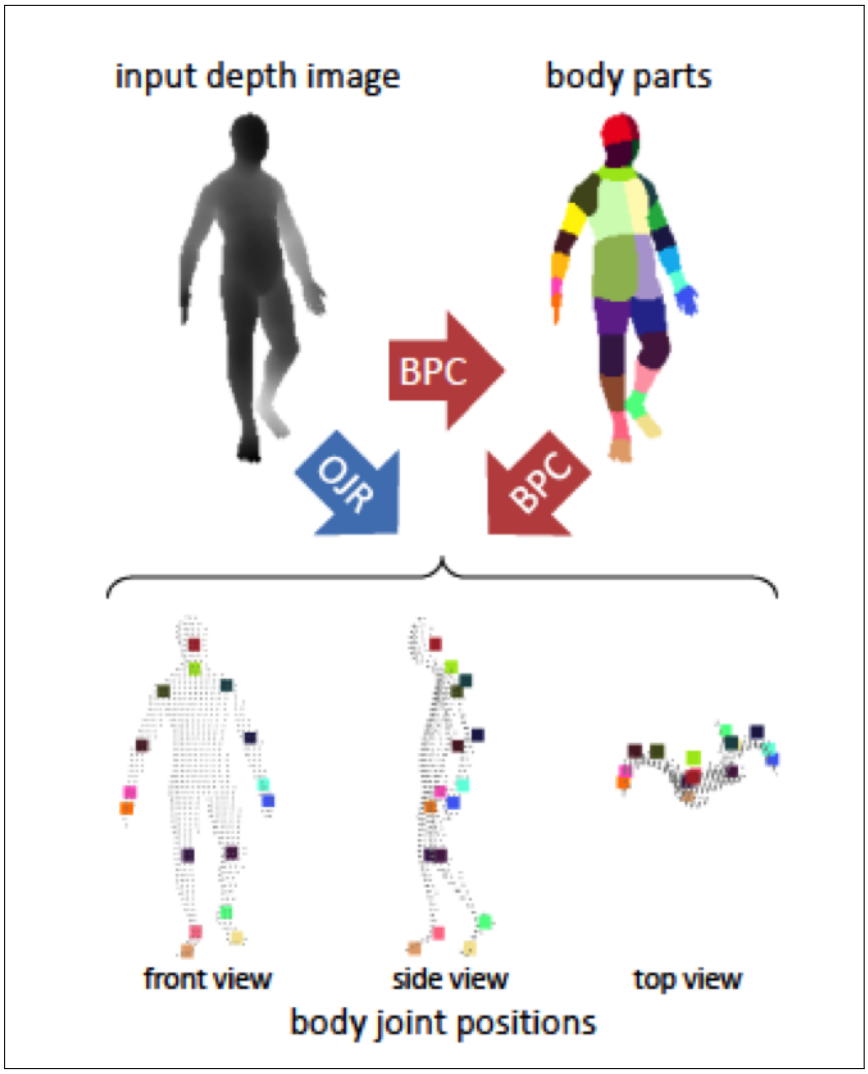

Figure 2.11: Overview of the two methods for predicting the positions of the body joints. Body part classification (BPC) has an intermediate step where the body parts are labeled before the location of the joints are found. Offset joint regression (OJR) directly estimates the joint positions using regression and a set of joint position predictions [36].

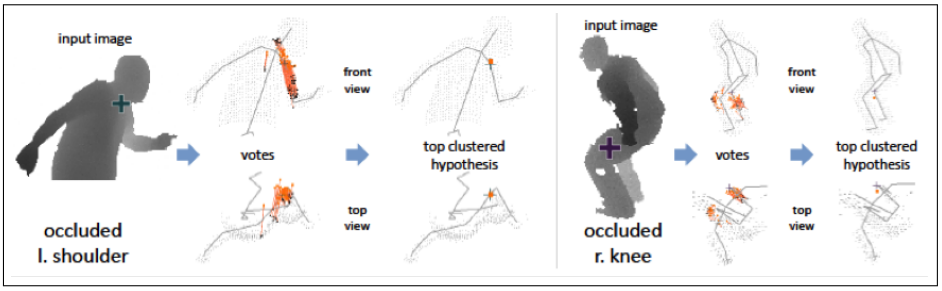

Figure 2.12: Offset joint regression uses regression from each pixel, illustrated as black squares here, to obtain a set of 3D votes for the joint position represented by the orange lines. The set of joint position predictions are aggregated using the mean shift algorithm and the prediction with the highest confidence is shown as the top clustered hypothesis [36].

Each pixel can have multiple votes for each body joint in the OJR algorithm, and achieves a more accurate body joint estimation even when the joint is occluded by other parts of the body. The BPC algorithm ended up with only one single vote for each body joint, which gives a higher uncertainty in the position of the joint. The OJR algorithm therefore seems to be able to solve the problem related to tracking body joints when the player is standing with its side towards the Kinect v1sensor. Microsoft promoted the Kinect v2 with improved skeletal tracking and there is reason to believe that the OJR algorithm is the background for the Kinect SDK 2.0 upgraded body tracking abilities.

Both algorithms are trained by using many real and synthetic depth images to avoid overfitting the data. Some of the base character models used for rendering the synthetic data are seen in figure 2.13. Several parameters were used to create a wide variety of 3D models that try to cover all possible body shapes, sizes, poses and orientations.

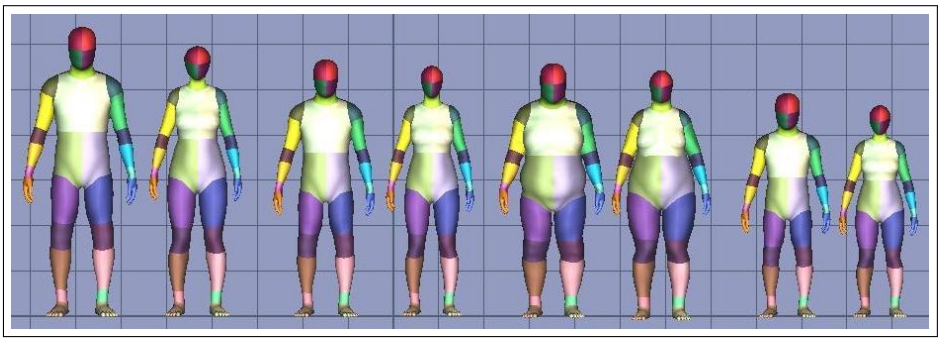

Figure 2.13: A few of the 15 base character models used for the body training and test data [36]

Real images for the purpose of training the algorithms were acquired using

motion capture where hundreds of thousands of frames were recorded of people performing all kinds of gaming and navigational poses. The synthetic and real images used in the training process are compared in figure 2.14 and the only visible difference is that the real images have high-frequency texture. Both sets of images have variations in pose, shape and clothing that are necessary in order to create a robust tracking algorithm [36]. There are also images where parts of the body is intentionally occluded or outside the sensor's field of view.

Figure 2.14: Synthetic and real images used as training data with varying poses, shapes and clothing. The black bodies are taken from depth images and next to each of them are the corresponding ground-truth labels for each body part [36].

2.2.3 Skeletal Joint Smoothing

Like most other systems that use a sensor to capture data that represents a physical quantity, the Kinect sensor is prone to errors and noise in the measurements. The skeletal joint positions that the Kinect SDK returns as raw data, have noise because of different factors such as the distance between the sensor and the player, the body size and shape of the player, and the lighting conditions in the room. As mentioned earlier the skeletal tracking cannot detect that a person is standing in a sideways pose and if this is the case, the skeletal joint data will either not be returned, or in most cases the joint positions recorded are erroneous.

According to Azimi and the *Skeletal Joint Smoothing White Paper* [5] the skeletal joint data are accurate but imprecise in relation to the real world joint positions. There is no systematic measurement error in the joint position data and therefore the positions are expected to be nearby the real world joint positions with an accuracy range of 1 centimeter. Even though the system is accurate in the joint position data there is a lack of precision which means the joint positions are not always correctly identified in every frame. This form of noise is a small factor in the joint position data, but noise due to inaccuracy can also occur when the system does not have enough information to find the location of the joint. Typical situations where this is the case is when some body parts are not in the sensors field of view, or occluded by other objects or body parts. The skeletal tracking system is then able to infer positions of the joints in question and convey the low confidence in determining the specific location of the joint in the current frame by returning the tracking state of the

joint(s) as inferred. The inferred joint positions may have a bias in addition to a drastic increase in the noise level. This can be clearly seen when the player is standing with the side facing the sensor and the system attempts to place skeleton joints in extremely unrealistic locations.

In his paper Azimi gives a presentation of techniques that will filter the skeleton joint data for use in applications for the Kinect sensor. Smoothing the joint position data will increase the latency to some degree dependent on the type of filter used. Figure 2.15 show a plot of the input and output from applying a smoothing filter to a joint movement. The latency is recognized as the lag between the input and output graphs, and the magnitude of lag depends on the speed of the moving joint.

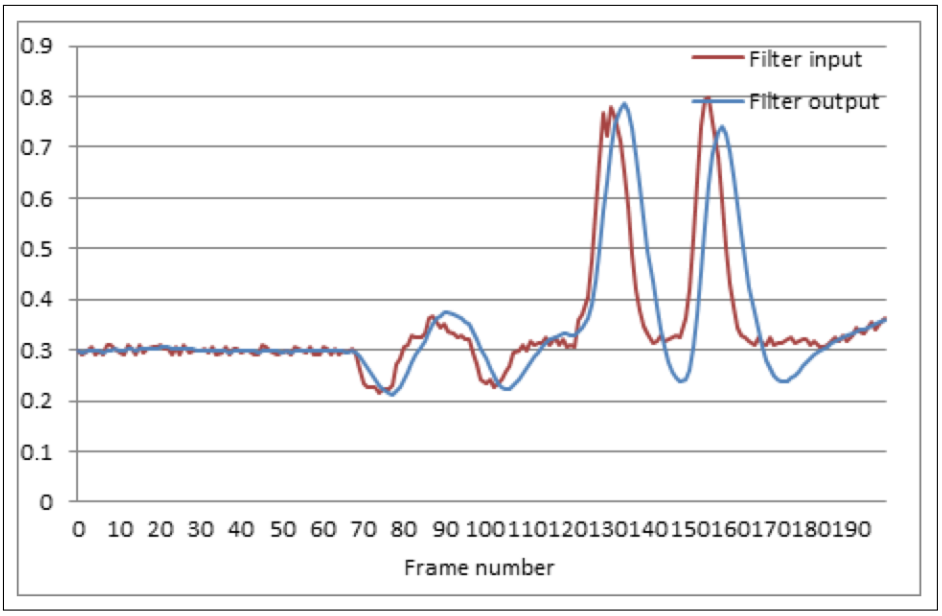

Figure 2.15: Plots of the input and output from applying a smoothing filter to joint movement [5].

To avoid lag joint positions in subsequent frames can be predicted by including forecasting in the smoothing filter. This method estimates the joint positions based on data from previous frames and reduces the delay, but is also less accurate if the motion performed is unpredictable in nature. Different body parts may move faster than others in certain movements and require a particular filtering technique to optimize the result.

In the application chosen for the Kinect in this thesis we decided to use the raw joint position data in the calculations of the different biomechanical measurement. One can argue that the results could benefit from smoothing, but since a big part of the goal of the experiments was to evaluate and compare the data to a different system, the decision was made to use the skeletal tracking out of the box. The post-processing of the data from the Qualisys marker based

system, includes smoothing the marker-position data by applying Woltring's generalized, cross-validatory spline smoothing. Spline functions are commonly used to smooth and differentiate data in biomechanics since they resemble human motion and can eliminate random noise that is normally distributed [6]. Woltring's method finds the optimal smoothing level based on cross-validation to avoid over- and under-smoothing which are results caused by user defined parameters [43].

2.3 Joint Kinematics in Three Dimensions

The marker based systems create 3D models of motion using marker positions that are strategically placed according to the anatomical features of the human body. These marker positions are used in the calculations of local joint coordinate systems that describe bone segment movements as accurately as possible without using an invasive method. The kinematic joint movement in three dimensions can be described geometrically using 6 parameters using the convention presented by Grood in a paper from 1983 [18].

2.3.1 Joint Coordinate System

The Joint Coordinate System (JCS) described in the paper by Grood [18] has 6 components that give a clinical description of the motion between two rigid bodies in terms of rotation and translation. One of the advantages of this coordinate system is that it is not necessary to specify the order of translations and rotations that is involved for a joint movement. To express the relative position between two rigid bodies a Cartesian coordinate system is specified for each of them. In the drawing in figure 2.16 on the following page these coordinate systems are illustrated with their respective origins at O_A and O_B.

The angular position between the two bodies can be described with three independent angles that correspond to rotations about three spatial axes. The unit vectors for two of the axes are shown in figure 2.16 on the next page as e_1 and e_3. They are both fixed to each of the bodies unlike the axis called **F**, which is a floating axis that is perpendicular and moves in relation to the other body-fixed axes. The unit base vector for the floating axis is shown as e_2. A body's rotation about its own fixed axis is measured in relation to the other body by the angles α and γ between the floating axis and the reference axis called $e_n^r, n = 1, 3$ in figure 2.16 on the following page. The magnitude of the rotation about the floating axis is the angle between the two fixed axes; β. We now have three angles that represent the relative angular position between two rigid bodies, more generally known as Euler angles. The line from point P_A to point P_B represents the displacement vector and translation of the joint. The reference points are chosen with specific applications in mind, and this determines the magnitude and components of the translation vector. Knee joint motion is of particular interest when the goal is to determine Anterior Cruciate Ligament (ACL) injury risk. The ligament is located on the front side of the knee and ACL-injuries are common during non-contact movement in sports such as handball, basketball and soccer.

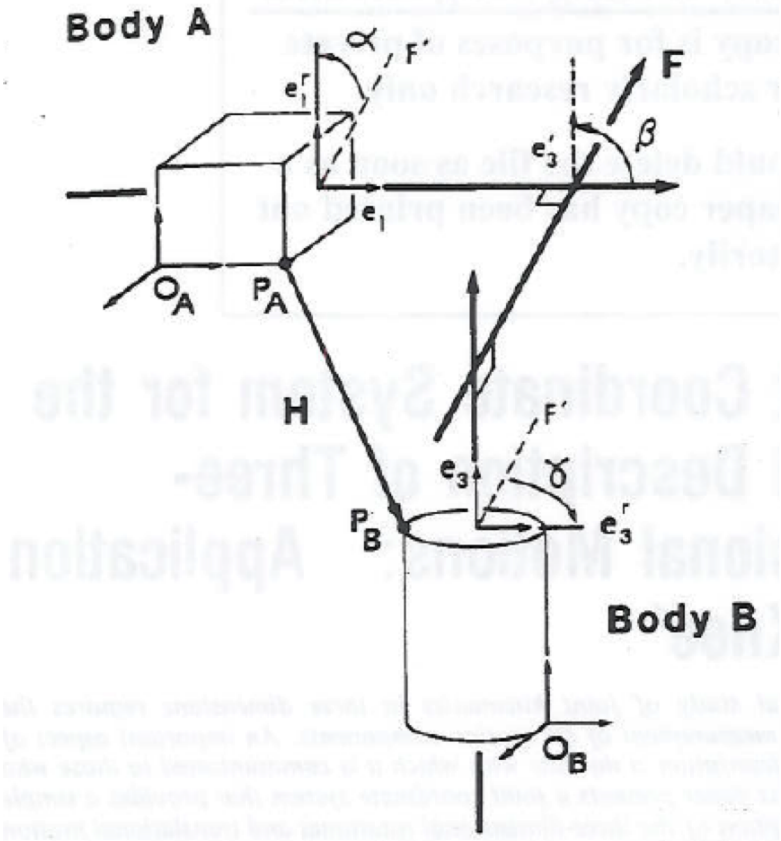

Figure 2.16: The relative motion between body A and body B can be described using the generalized joint coordinate system with three axes. The unit base vectors for the two body fixed axes are marked as e_1 and e_3, and the third floating axis **F** is perpendicular to them both.The unit base vector for the floating axis **F** is e_2 [18].

Coordinate System for the Knee Joint

The joint coordinate system (JCS) for the knee is illustrated in figure 2.17, but before these three axes are defined as previously mentioned, a Cartesian coordinate system is assigned to each bone segment to provide a description and orientation of the segment.

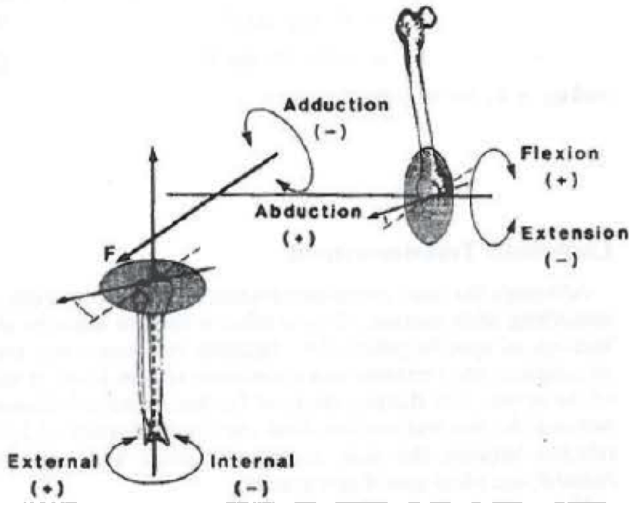

Figure 2.17: Joint coordinate system for the knee with three axes and the rotations that define the joint angles. Knee flexion and extension is about the femoral fixed axis while the external and internal rotation is about the tibial fixed axis. Adduction and abduction is about the floating axis **F** [18].

In this case we have the tibial and femoral bones seen in figure 2.18 on the following page with their respective coordinate system, where the z-axis is positive in the proximal direction (toward the center of the body). The y-axis is positive in the anterior(frontal) direction and the x-axis is positive to the right [18]. It is an advantage to define the Cartesian coordinate system so that two of the axes correspond to the body fixed axis and reference axis. It is also natural to put the origin of the Cartesian coordinate system the same place as the points used to find the relative translation between the bodies. Looking at figure 2.17 and the JCS for the knee, we see that the axes are conveniently chosen so that the measurements of interest are found as rotations about the body fixed axes.Knee flexion and extension is about the femoral fixed axis and the external and internal rotation is about the tibial fixed axis. Adduction and abduction is measured as rotations about the floating axis **F** [18].

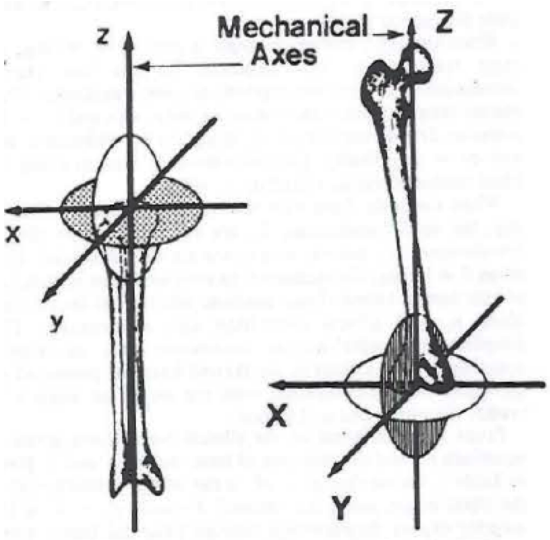

Figure 2.18: Cartesian coordinate system for the tibial (left); x, y, z and the femoral (right); X, Y, Z bone. The z-axis is positive in the proximal direction toward the center of the body. The y-axis is positive in the anterior (frontal) direction and the x-axis is positive to the right [18].

2.4 Assessment of Human Motion

Injuries to the Anterior Cruciate Ligament (ACL) are very common among athletes in various sports that require movement in all directions and angles. Tearing of the ACL is a severe season-ending injury that usually requires surgery, and puts the injured out of play for about six months. ACL injuries occur at a much higher rate in female athletes than in male, and studies conclude that this is related to knee stability and control during sports-related movement [20].

Neuromuscular control is important to be able to handle the force/torque that the knee joint experiences during for example jumping and landing. Many studies aim to find the reason why female athletes have a higher risk of obtaining knee-injuries, and if there are any indicators that can be used to classify individuals that are at high risk for ACL injury.

A study based on video analysis of twenty recorded incidents of ACL-injuries, and interviews of injured female team handball players, revealed that "a forceful valgus collapse with the knee close to full extension combined with external or internal rotation of the tibia" was the main injury mechanisms for ACL injuries [33]. The assumption that some individuals have a predisposition for ACL injuries is supported by the fact that they mostly occur in non-contact situations.

2.4.1 Tasks for Identification of Athletes with High Knee Motion

Some biomechanical measures have been established as predictors for ACL injury risk, and screening sessions are meant to evaluate these measures in sports-specific situations. Screening for ACL injury risk implies analyzing knee stability and control during sport-specific tasks. The tasks mimic movements with the same characteristics known to have caused ACL injuries. The vertical drop jump and single-leg squat are usually performed in screening test to provoke the movement that reveals if an athlete has high knee motion. These tasks are used in studies that aim to establish indicators of knee stability and ACL injury risk, and are therefore also chosen for the validating experiments described in the following chapter of this thesis.

Vertical Drop Jump

The vertical drop jump (VDJ) is a preferred task used to identify poor knee stability which is a risk factor for ACL-injury [20]. The VDJ is performed by having the athlete stand on a platform about 30 cm above the ground, before dropping down to the ground. Once the feet hit the floor the athlete is instructed to perform a maximum vertical jump. During the performance of a drop vertical jump test, there are especially two critical points that are important for computing the biomechanical measurements of interest. The first is at the athlete's initial point of contact (IC) with the floor after falling from the platform, and the other is at the point of peak knee flexion (PF) during the contact phase. The contact phase is defined as the period in time from IC until the feet has left the floor for the vertical jump.

Single-leg Squat

Most athletes that injure the ACL do so in situations where one leg is taking the majority of the load [21]. Single-leg squat is a used to evaluate knee stability without the support the other leg would give in a two-legged test [39]. This task is related to several movements that are common in sports and activities, and is useful to assess differences in knee control in each leg. A study by Aageberg concluded that a single-leg squat is a reliable task in assessment of medio-lateral motion of the knee [1].

2.4.2 Biomechanical Measures and Kinematics in Sports

Sports biomechanics is the field of science where mechanical laws are used to describe human sports-related movement, and provides a way of understanding sports injury mechanisms as well as athletic performance. The relationship between the forces and the human motion caused by the forces, is known as kinetics. In this thesis we are only concerned with the mechanics of the human movement without looking at the forces that cause the motion. Kinematics, or "geometry of motion", is the official name for this branch of sports biomechanics which describes human movement using geometrical terms. An example of the coordinate system and joint positions used as basis for different biomechanical measures [28], is illustrated in figure 2.19 on the following page.

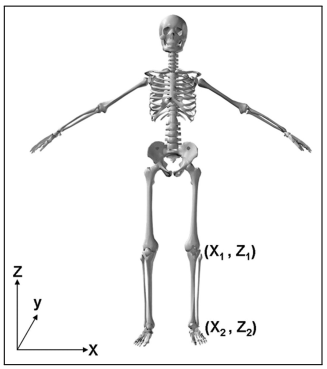

Figure 2.19: Illustration of coordinate system and joint positions from study by Myer et al. [28]

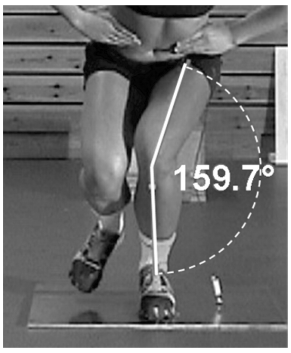

Figure 2.20: Frontal plane projection angle (FPPA) during single-leg squat [39].

Knee Valgus Motion

Excessive knee valgus motion or "dynamic valgus" shown in figure 2.21 on the next page, is "the medial collapse of the knee during dynamic tasks" [32], and is an indicator of ACL injury risk. Dynamic Valgus was defined by Hewett et al. [20]as "the position or motion, measured in 3 dimensions, of the distal femur toward and distal tibia away from the midline of the body. Dynamic valgus may have included the indicated motions and moments." Dynamic knee valgus movement occurs when the knee abduction angle and external abduction moment is high, which frequently occurs in female athletes when performing sports that require cutting, jumping and landing such as basketball, handball and soccer. The characteristics of the valgus knee motion is shown in figure 2.22 next to the normal and the vargus knee alignments.

$$\text{Knee valgus motion} = |x_1(IC) - x_1(PF)| \tag{2.1}$$

Knee-to-ankle Separation Ratio (KASR)

The knee-to-ankle separation ratio (KASR) is a measure that can be estimated from two dimensional video and is obtained by dividing the distance between the knee joints by the distance between the ankles. This measure requires a two-legged task such as the VDJ for the value to make any sense.

$$\text{Knee-to-ankle ratio} = \frac{|x_{1,right} - x_{1,left}|}{|x_{2,right} - x_{2,left}|} \tag{2.2}$$

Frontal Plane Projection Angle (FPPA)

The angle between the thigh and the leg is recognized as the frontal plane projection angle (FPPA), and is calculated in the frontal plane as the angle between the vector from the knee joint down to the ground (normal to the horizontal

[1]Illustration from: karategirl.org

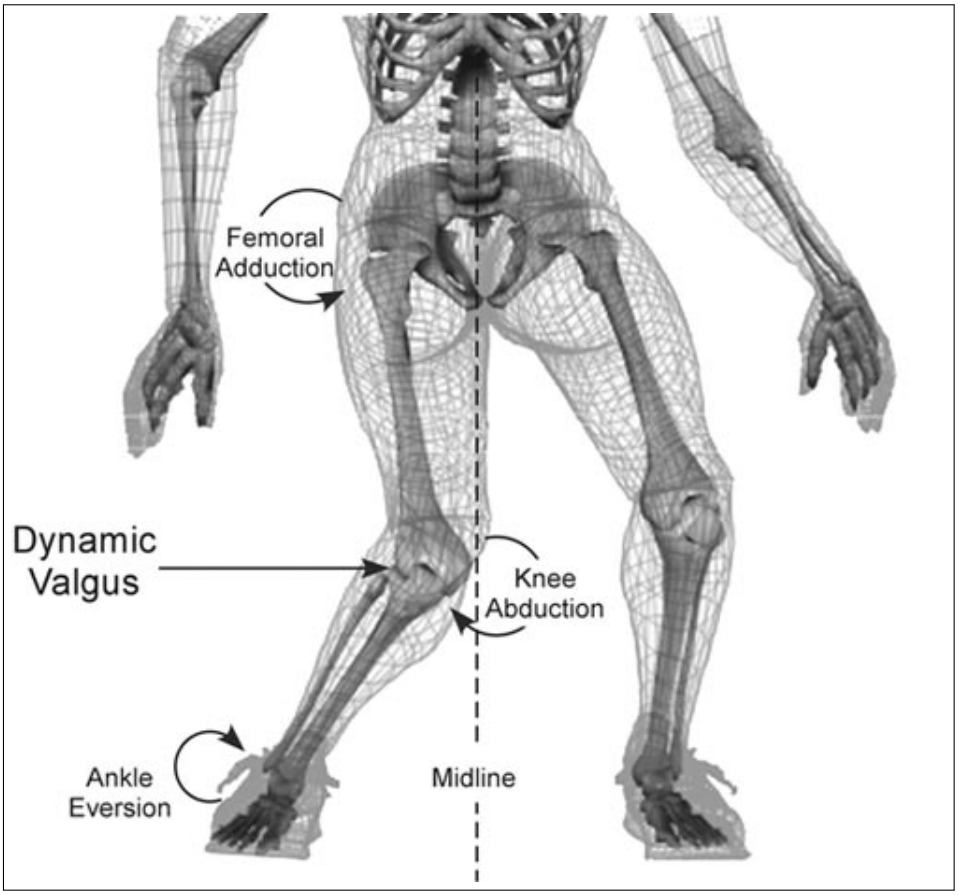

Figure 2.21: Dynamic Valgus was defined by Hewett et al. [20]as "the position or motion, measured in 3 dimensions, of the distal femur toward and distal tibia away from the mid line of the body. Dynamic valgus may have included the indicated motions and moments."

axis), and the vector from the knee joint to down to the ankle joint [40]. The image in figure 2.20 on the preceding page shows a measurement example of the frontal plane knee angle during a single-leg squat. In the drop vertical jump task, this variable is usually measured at two different times during the performance of the task; at the athlete's initial time of contact with the floor after falling from the platform, and at the time of the maximum knee flexion (peak flexion) after landing right before leaving the floor to perform the vertical jump.

$$\vec{KA}_x = \frac{\langle x_2 - x_1, z_2 - z_1 \rangle}{\|\langle x_2 - x_1, z_2 - z_1 \rangle\|}$$
$$\vec{KG}_x = \frac{\langle 0, z_2 - z_1 \rangle}{\|\langle 0, z_2 - z_1 \rangle\|}$$
$$\text{Frontal plane knee angle} = \cos^{-1}(\vec{KA}_x \cdot \vec{KG}_x)$$

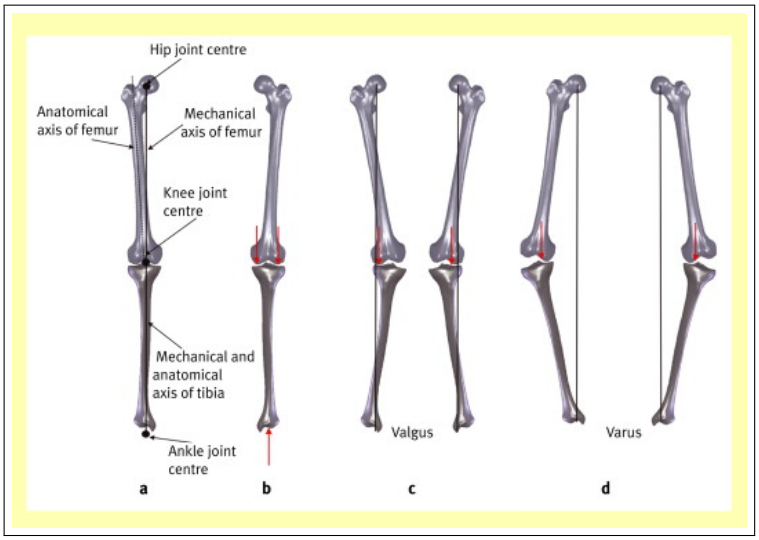

Figure 2.22: The anatomical and mechanical axes of the leg is shown to the left (*a*). The neutral knee alignment is illustrated in *b* and the knee valgus ("knock-knee") alignment is illustrated in the center (*c*. The knee varus alignment ("bow-legged") is shown in *d*.[1]

Knee Abduction Moment

High knee abduction moment during a VDJ is an indicator of ACL injury risk in female athletes [20]. This measurement is related to the forces (kinetics) that cause the dynamic knee motion during sport movement. The two factors that contribute to the maximum knee abduction moment are ground reaction force (GRF), and the GRF moment arm in the frontal plane. Marker based motion analysis are performed in a laboratory setting that has force plates built into the floor that can measure the ground reaction force for each leg during the contact phase of a drop jump. The ground reaction force is used in the calculation of the knee abduction moment. Reducing the knee abduction moment is likely to reduce the risk of ACL injury, and a simplified method for calculating the knee abduction moment that ignores the inertia of the bone segments has been presented by Kristianslund [22]. It is not possible to calculate moment with a MLS such as the Kinect, that does not include the use of force plates. In a clinical settings there is no known way of easily acquiring forces, so this measure was not relevant for our experiments.

2.4.3 Injured versus Non-injured

In a study done by Hewitt [20] to identify predictors for ACL injury risk, more than 200 female athletes from different sports were screened using a 3D motion capture system. The vertical drop jump (VDJ; explained in section 2.4.1 on page 38) was the task used for the screening tests in the study by Hewett. Variables were recorded during the performance of the VDJ and used to calculate the knee valgus motion (knee abduction angles).

Hewett et al. [20] conducted the study before the respective seasons, and a total of 9 test-subjects ended up tearing their ACL. Analyses of the testing data showed that all of the injured athletes had greater knee abduction angles during the vertical drop jump task, than the subjects that remained uninjured. A biomechanical model of the mean knee joint kinematics during the drop vertical jump at different points during the task, can be seen in figure 2.23. The conclusion was that the increased knee valgus motion observed in the female athletes that later tore their ACL, indicated higher risk of injury.

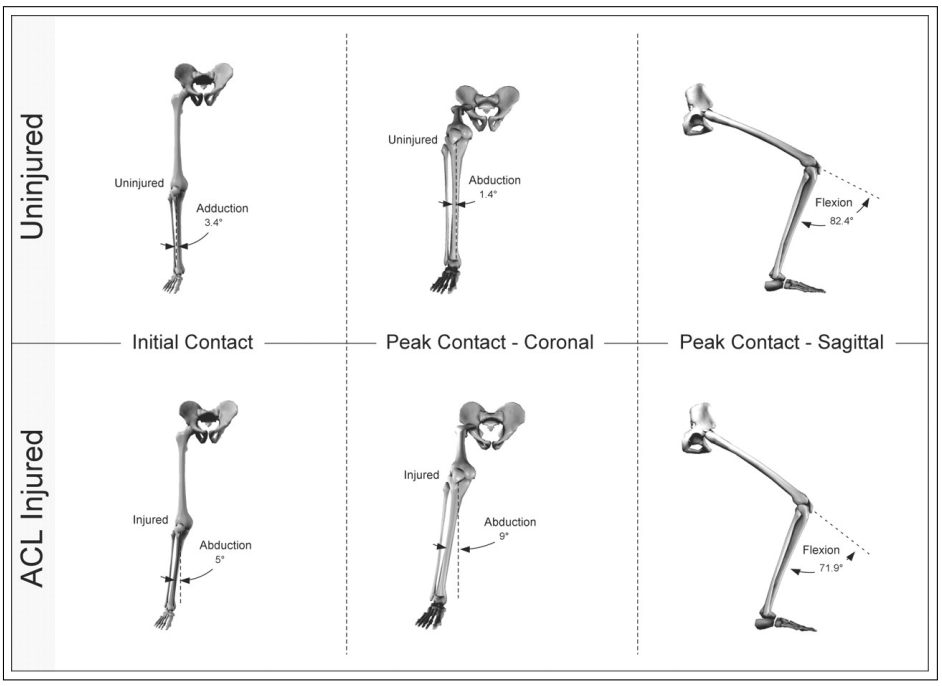

Figure 2.23: Biomechanical model of mean knee joint kinematics during VDJ task; injured vs. non-injured [20].

2.4.4 Previous Work Using Markerless Motion Capture Systems

The range of motion between a marker based system and a markerless system (Microsoft Kinect) was compared recently in a study by Bonnechère et al. [2]. The results from the comparison by Bonnechère showed poor to no good agreement between the two systems in the measurements of knee flexion. Most likely the discrepancies were caused by inaccurate estimation of the joint centers which was done frame-by-frame. This is not dependent on the motion performed and the differences between the systems could probably be reduced with an improved estimation of the joint centers or better synchronization. Another limitation to this study is the version of the software used; Microsoft Kinect SDK v.1.5 which is now outdated and a newer release could have provided better results in terms of validity and reliability.

A recent study by Stone [40] evaluated the ability of the Kinect sensor to screen for ACL injury. The measurements were calculated for the frontal plane only because not enough information about the bone segments was acquired from the Kinect. The study concluded that there was an agreement between the measurements from the two systems based on high intra-class correlation coefficients (ICC) for the Knee Valgus Motion, knee-to-ankle separation ratio and frontal plane projection angle(fppa) at initial contact(IC) and point of peak flexion(PF).

2.4.5 Methods for Screening for ACL Injury Risk

As mentioned earlier there is a need for methods that have the same reliability as marker based systems but that are easier to implement. A few other techniques that are comparable to the markerless approach, are observational screening and two-dimensional video analysis.

Real-time Observational Screening

There are few real-time methods for screening athletes for knee control and stability, but in clinical settings observational screening is often used. This assessment method requires visual inspection and evaluation by physiotherapists that classify knee control during specific tasks. This rating method can be compared to two dimensional video analysis, but as explained by Stensrud [39] it is a subjective method and depends on the raters experience with screening athletes for knee control. A study done by Nilstad [32] used real-time observational screening by trained physiotherapists to assess knee control of female athletes during a vertical drop jump. The visual inspection of frontal plane knee control during the landing phase of a vertical drop jump was compared to the knee valgus angles and abduction moments calculated from three-dimensional motion analysis. The study concluded that athletes with high knee valgus angles during a drop jump landing can be identified by physiotherapists using real-time observational screening.

2D Video-based Techniques

Video-based techniques that record data in two dimensions using standard video cameras, rely on variables in the frontal plane to measure knee motion. In a study by Myer [29] the knee separation distance and frontal plane projection angle was measured in two-dimensional video recordings, and then compared to the same measures obtained from 3D motion analysis. The correlation coefficients showed a high intraclass agreement and proves that video analysis can give a more standardized and detailed assessment of knee control, than real-time observational screening such as the study by Nilstad [32]. Mizner et al.[26] compared two-dimensional (2D) measurement techniques to a three-dimensional motion analysis system to determine the association between the variables from the two different systems. The knee-to-ankle separation ratio was extracted from the 2D video in Mizner's study by defining the lines between the knee joint centers and ankle joint centers in the frontal plane. The vectors used in the calculation of KASR in the frontal plane are illustrated in figure 2.24. An ACL prediction algorithm validated by Myer [28] in 2011 identifies female

athletes with high knee abduction moment, which implies risk of ACL injury. The algorithm is based on measures obtained from clinical trials and include knee valgus motion, knee flexion range of motion, body mass, tibia length and quadriceps-to-hamstrings ratio.

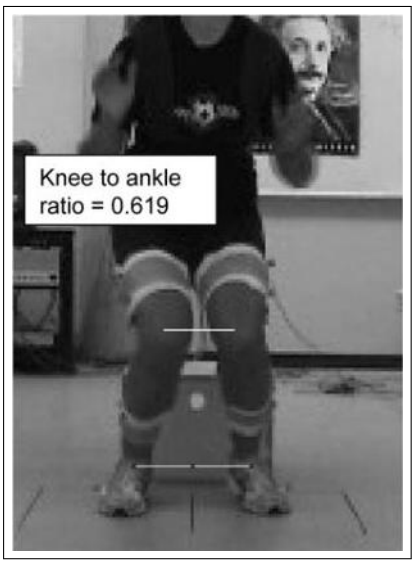

Figure 2.24: Illustration of the vectors used in the calculation of the knee-to-ankle separation ratio (KASR) in the frontal plane [26].

Chapter 3

Material and Methods

This chapter contains a detailed description of the method we chose for validating a markerless motion capture system (MLS) against a marker based system (MBS). The research questions mentioned in the introduction are related to ACL injury risk screening based on analyses of motion patterns of athletes performing specific tasks. The tasks and collection of data are therefore chosen with that in mind, but it was also of interest to evaluate the general tracking abilities of the MLS to see how this affects the calculation of the kinematic measurements.

The study was performed in cooperation with the Oslo Sports Trauma Research Center (OSTRC) at the Norwegian School of Sport Sciences, and their motion analysis laboratory was used for the experiments. The MLS used in the experiments was the Microsoft Kinect sensor (version 1 and 2), and the MBS was integrated in the lab consisting of sixteen 480 Hz infrared cameras (ProReflex, Qualisys, Gothenburg, Sweden). The lab also has two 960 Hz force platforms (AMTI, Watertown, Massachusetts, USA) that were not used in the experiment. The experiment was set up in a similar testing-environment that has been used by the research group at the OSTRC to evaluate risk factors for ACL-injuries in elite female soccer players [12].

This study was carried out as 3D motion analyses with screening sessions to determine the accuracy and tracking ability of MLS. We wanted to analyze these specific movement using discrete variables:

1. Knee motion during vertical drop jump (VDJ)
2. Knee motion during single-leg squat (SLS)
3. Knee motion during front-oriented squats
4. Knee and hip motion in the sagittal plane during side-oriented squats

3.1 Subjects

The test subjects, that were recruited for this study, were healthy female athletes that train on a regular basis. The subjects signed an informed consent form

that also permitted me to use photos taken during the testing in my thesis. The subjects that were under 18 years old had to get permission from their parents or legal guardian to participate in the study.

The test subjects for the first session of testing were twelve female basketball players between the age of 16 to 32 from different teams in the top league in Norway.

The second session was a pilot-testing session with the Kinect v2 sensor with three of the subjects that also participated in the first testing session.

3.2 Instrumentation

The lab setup was similar for all sessions and the cameras were positioned in 360 degrees around the recording area. The Kinect sensor was placed on top of a table and taped in a rigid position to make sure it did not move for the duration of the testing. The distance between the sensor and the test subject was between two and four meter, and during the testing the subjects were facing the sensor (except for one of the tasks). It was important that the subject was within the field-of-view (FOV) of the Kinect sensor at all times, and also outside of the FOV of the MBS infrared cameras to avoid possible interference during the calibration.

Kinect Studio was connected to a Kinect-enabled application for recording and saving of the Kinect Studio files for each subject and trial. The format of the files from Kinect Studio (*.xed* for Kinect v1 and *.xef* for Kinect v2) is not publicly available, so the recordings had to be processed through an application that could save the infrared image frames and raw 3D data. Kinect Studio 2.0 does not support RGB image stream at this time.

The MBS was controlled by a OSTRC employee who also assisted on the other parts of the testing.

3.3 Protocol

A physiotherapist was trained to identify and place 37 reflective markers on anatomical landmarks on each of the test subjects for session 1 and 2. A model of the athlete and the placement of the markers is illustrated in figure 3.1 on the next page. The physiotherapist also made sure the tasks were performed correctly according to the guidelines in the protocol for the testing, which is explained later. In some cases the markers fell off during the testing and had to be reattached by the physiotherapist.

After the marker placement was completed, the test subject began warming up to prepare for the testing. The warm-up was intended to ensure that the subjects gave maximum effort in all trials of the tasks in the actual testing. The exercises chosen for the warm-up are listed in table 3.1 on the following page.

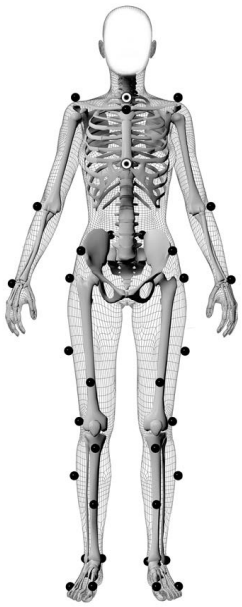

Figure 3.1: Model of athlete with anatomical landmarks for placement of the markers

Table 3.1: List of warm-up exercises for test preparation.

Warm-up		
Exercise	**# of reps**	**weight**
Two-legged squat	10	0 kg
	10	20 kg
	5	50% of body weight
Jumping squats	10	None
Single-leg squat	5 each leg	None

3.3.1 Static Trial

The MBS (Qualisys) requires a static trial for each subject before the actual testing begins. For the static trial the subject stood for ten seconds in the middle of the testing area in the anatomical position shown in figure 3.2 on the next page. The static trial defines the anatomical coordinate systems for the subject and the static (neutral) position of each marker. During the tasks the tracked marker positions in the Qualisys system are relative to the positions in the static trial. The static trial was also recorded by the MLS (Kinect), but for most of the subjects the sensor was unable to locate the joint positions because the subject was standing completely still. In order for the Kinect application to locate the joints in the static position, the recording had to be started before the subject had raised their arms. Even though the application seemed to track the skeleton during the record-time, the processing after each session revealed that the skeleton data was not available. The photo in figure 3.2 show the Kinect

application during recording of the static trial for one of the subjects, where the skeleton and joint positions appear to be tracked.

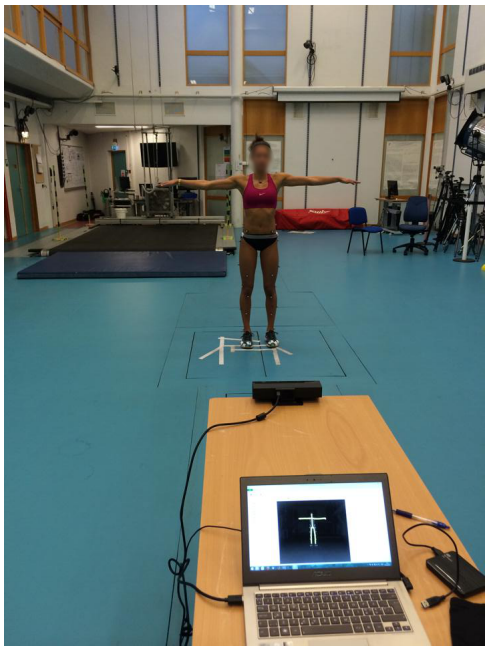

Figure 3.2: Photo of a subject standing in the anatomical position during the static trial. The Kinect v2 sensor is placed on the table in front of the subject and the Kinect Studio 2.0 recording tracks the skeleton.Although not clearly visible in the image, the subject has 37 markers placed on anatomical landmarks on her body.

Tasks for Evaluating Knee Motion

In section 2.4.1 on page 38 two tasks were described as frequently being used in evaluation of knee stability and screening for ACL injury risk. The vertical drop jump (VDJ) and the single-leg squat (SLS) were therefore the obvious chose of tasks to implement in our testing. In addition to these two tasks, one other task was included to further test the ability of the Kinect to track sports-related movement. This task was the two-legged squat with bar which is a key strength exercise used in training for a wide variety of sports, including basketball. The choice of this task was also motivated by the objective to evaluate the side-tracking ability of the Kinect (MLS) sensor. The knee and hip angles are of interest as these often are indicators of whether the technique is correct or not. Squats should be performed with minimal bending of the upper body to avoid potential back injuries induced by the exercise. This becomes more and more important as the weight on the bar increases. The squatting technique is usually evaluated visually by trainers looking at the exercise being performed from the side. To simulate a realistic training setting, the two-legged squat was done with a bar and weights corresponding to approximately half of the body weight

of the subject. The squat was recorded from the front and from the side with reference to the location of the Kinect sensor. The

The subject started with the vertical drop jump (VDJ) task and repeated the motion for as many times it took to have 5 successful trials, recorded by both the markerless (Kinect) and the marker based system. A biomechanical illustration of VDJ is shown in figure 3.3. The next task was the two-legged squat with bar (front-view) in which the subject completed 5 consecutive repetitions descending as deep as possible on each rep. The procedure was the same for the next squat recording where the subject's side was oriented towards the Kinect sensor.

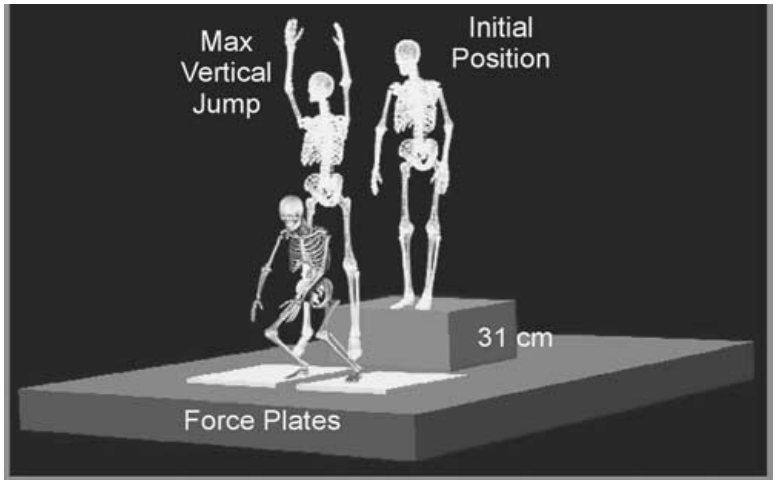

Figure 3.3: Biomechanical illustration of the vertical drop jump (VDJ) task [20]. (For our experiments the height of the box used was 30 cm, and the force plates were not activated.)

The single-leg squat was first performed for the left leg, and then for the right leg, each with 5 consecutive repetitions. Only the leg doing the squatting was allowed to touch the floor for the duration of the task. A piece of string with a metal-ring attached to the end was taped to side of the hip of the subject during the SLS task. The length of the string was adjusted so that it touched the floor at the time the subject had descended to an angle of 90 degrees between the thigh and the shank. This indicated that the subject was deep enough into the SLS position.

3.4 Data Processing

The OSTRC has a very specific procedure for producing output relevant for their research projects with the Qualisys marker based motion capture system (MBS). Since this thesis is based on a cooperation with OSTRC, it was important to us to follow the same procedures. Before any 3D data could be processed from the MBS, the Qualisys motion capture *.qtm* files had to be processed in the Qualisys Tracking Manager (QTM) application. The different markers trajectories had to be tracked according to a 3D model with labels for the 37 anatomical

landmarks. After the initial labeling of the marker positions, the 3D model was saved and trained to recognize the correct labels when applied to recordings of other subjects. A screenshot of the QTM application is seen in figure 3.4

In many cases the trajectories of some markers had gaps, and these gaps, except for the ASIS-markers (anterior superior iliac spine) were filled using polynomial interpolation inside the QTM application. The goal of 100% tracking coverage for all of the marker trajectories was achieved in most cases. The gaps in trajectories were usually caused by a marker not being visible to the cameras due to occlusion. The missing parts trajectories were mostly the ASIS-markers (on the front side of the hips); since they were sometimes hidden as the subject landed in the VDJ task, or at the deepest descent in the squat tasks. Once the tracking for all the trials for all subjects was completed, the 3D data was exported to *.tsv* files.

The 3D data from the MLS (Kinect) was not processed with gap-filling or correction of inferred joints.

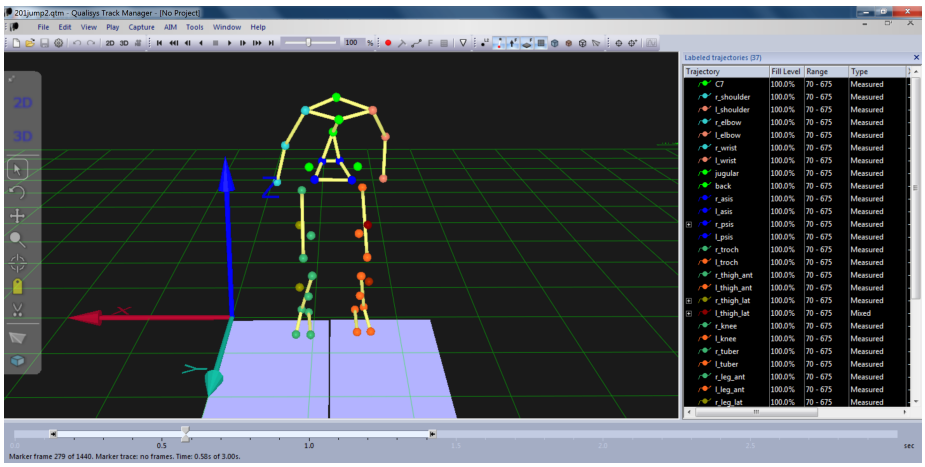

Figure 3.4: Screenshot of the Qualisys Track Manager software used to track the marker trajectories. The panel on the right side show the labels for the anatomical landmarks in the 3D model that points to the respective markers.

3.4.1 Markerless (MLS) Data Processing

The 3D motion capture data from the Kinect Studio 1.8 recordings was saved as binary files using the Kinect Stream Saver Application, developed by the AI & Robotics Research Team at the Toronto Rehabilitation Institute [14]. The binary files contained time stamps and 3D joint data for each recorded frame, and were imported into Matlab for further processing. Since the Kinect v2 sensor has not been on the market for very long, there was not any ready applications that could record the raw image and 3D data from Kinect v2. A Windows Presentation Foundation (WPF) application was developed to save the raw body joint data from the Kinect Studio 2.0 files (*.xef*). The application

was a modified version of the *Infrared Basics* application provided in the Kinect SDK 2.0. A screenshot of the WPF application, with a Kinect Studio 2.0 file as the input, is shown in figure 3.5.

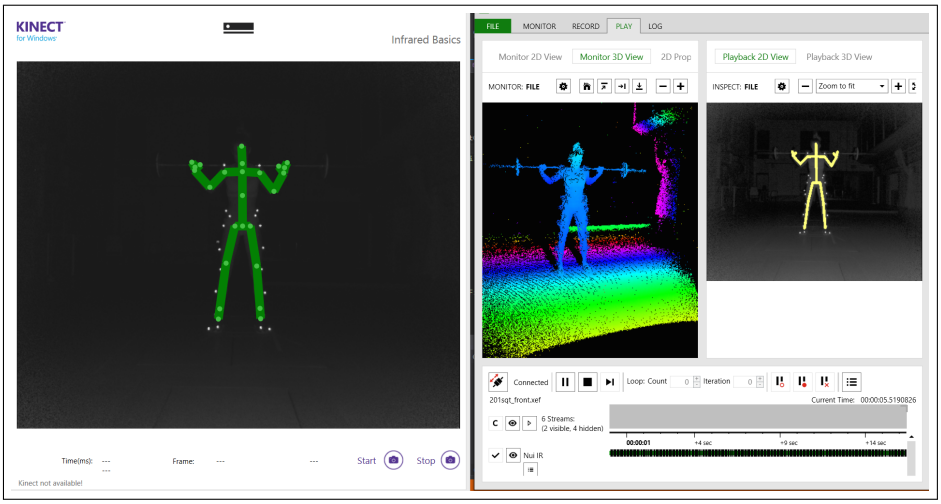

Figure 3.5: Screenshot of the Kinect-enabled WPF application (left) and Kinect Studio 2.0 (right) running on Windows 8.1. A Kinect Studio 2.0 file is played back and used as input to the application on the left that saves the raw 3D data.

The output from both sessions with the MLS included the 3D position and orientation for each of the body joints in the Kinect skeleton. The joint positions and orientations are given relative to the global coordinate system of the Kinect sensor.

The skeleton, or body which is the new official name in Kinect SDK 2.0, included five additional joint positions in addition to the 20 body joints from the previous SDK. The naming convention changed for some of the original joints. The extra joints are located in the tips of hands and feet of the skeleton. The two versions of the Kinect SDK skeletons and joints, relative to the human body, are compared side by side in figures 3.6 and 3.7 on the following page.

The orientation of the joint is given as a quaternion[1], where the basis is defined with the y-axis in the direction of the bone (using the knee as an example; Y lies on the line going from the knee to the ankle pointing away from the knee). The x-axis is perpendicular to the bone and the z-axis which is normal to the bone (y-axis). In the latest version of the SDK (2.0) the quaternion for each joint expresses the rotation from the global coordinate system for the Kinect sensor to the current orientation of the joint [31].

[1]A quaternion is a mathematical representation of an orientation using four numbers; a scalar component (w) and a 3D vector (x,y,z).

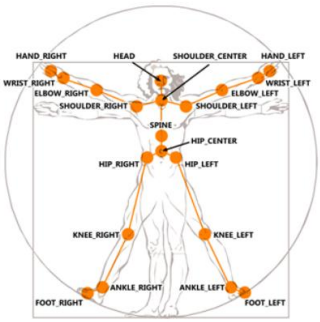

Figure 3.6: Joint positions (20) relative to the human body in Kinect SDK 1.8 [31].

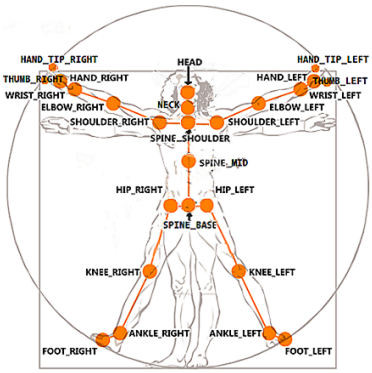

Figure 3.7: Joint positions (25) relative to the human body in Kinect SDK 2.0 [31].

3.4.2 Marker based (MBS) Kinematic Data Processing

The program that OSTRC uses to get the kinematic data and measurements from the Qualisys MBS, is implemented in Matlab. The *.tsv*-files from the exported *.qtm*-files are read into variables. Originally the program relies on the force-platforms, in processing of the vertical drop jumps, to determine the initial point of contact (IC). A slightly modified version manually specified the IC based on the vertical position of one of the toe markers. Some other adjustments had to be made in order to use the same program for this thesis. Access was given to parts of the source-code of the Matlab program in question, and inspired the development of a program for processing the data from the Kinect MLS in the same way.

As mentioned in the previous chapter in section 2.3 on page 34, the marker based 3D motion analysis uses the Joint Coordinate System (JCS) which provides a precise definition of the rotations of the bone segments in terms of clinical descriptions. This convention is common in the field of sports biomechanics since it is easily interpreted in terms of anatomical movement. Knee flexion/extension and abduction/adduction angles, as well as translational motion are examples of these movements. The coordinate system is comprised of both the proximal and distal bone segment; and, the rotations are defined as Cardan angles about the three axes. The details for the calculations and mathematical definitions for describing the joint rotations and translations will not be provided. It is sufficient to understand that JCS gives us an intuitive way of finding the kinematic measurements of interest in terms of human motion. The order of rotations for Cardan angles, using the leg and its segments as an example, as following:

1. Rotation about the flexion axis of the thigh(femur), then
2. Rotation about the "floating axis" (cross product of flexion axis of the thigh and long axis of the shank) and finally
3. Rotation about the long axis of the shank (tibia)

An illustration of the JCS for the knee as a four-link kinematic chain is inserted in figure 3.8. The relative motion between the segments can be described using a method commonly used in robotics, known as the Denavit-Hartenberg method [18].

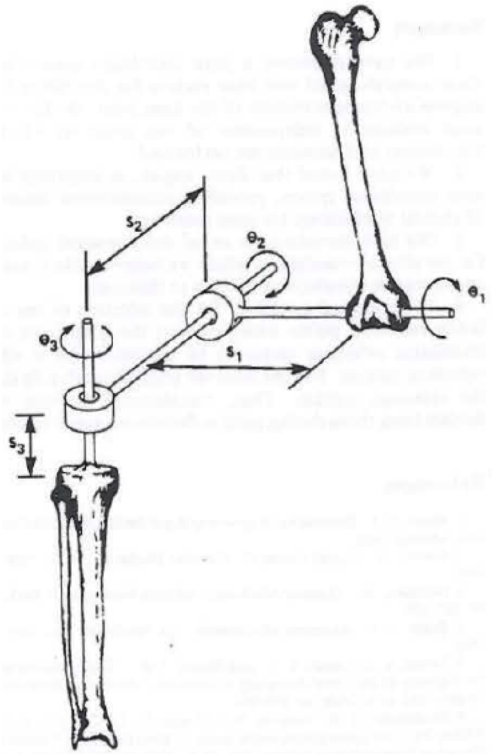

Figure 3.8: The Joint Coordinate System as a four-link kinematic chain illustrated with the tibial and femoral bones and cylindrical joints [18].

In the program for kinematic calculations for the MBS, the coordinate system (JCS) for the different joints are based on the "neutral positions" of the static trial. Søderkvist's algorithm is used to find the reference frame, described in section 2.3 on page 34, for the thigh and the shank. The algorithm creates a matrix of the positions of the markers (anatomical landmarks), and uses singular value decomposition of that matrix to determine the motion of the bone segments [38].

The MBS output was in the form of *.mat*-files with the dynamic marker-positions and plots of ankle, knee and hip kinematics (flexion, adduction and internal rotation). In addition, the joint center locations for the knee, ankle and hip were estimated and included in the output. The center of the ankle joint is estimated using Eng and Winter's method [15], where the joint center is defined to be 1 cm distally from the ankle marker. The hip joint centers are estimated based on an article by Bell et al. that compared different methods for predicting hip center locations [7]. The knee joint center location is determined

using the estimation method by Davis et al. described in an article on a gait analysis data collection technique in 1991 [13].

In this thesis we will distinguish between the knee flexion angle from the MBS that is calculated using JCS and Cardan angles, and the knee flexion angle between two vectors in 3D space using three joint center positions from the MLS; hip, knee and ankle.

3.4.3 JCS for Kinect Body Joints

An attempt was made to process the 3D data from the MLS (Kinect) to define the reference frames for the bone segments in the same way it is done in the MBS program. We hypothesized that the rotation matrices for each bone (joint) segment for the Kinect should be similar to the rotation matrices for the joint positions in the MBS data. The goal was to derive a method for extracting the JCS and Cardan angles for the Kinect data, using the joint positions and the corresponding joint orientations (quaternions) from the static trial. The motivation behind this was that if we found the JCS and cardan angles for the bone segments from the MLS data, we could use the same method as the MBS to calculate the knee kinematics shown in the plots in figure 3.9 on the next page.

The quaternions for the relevant Kinect body joints (hip and knee) were easily transformed into rotation matrices that corresponded to the rotation matrices for thigh and shank found from the MBS static trial. A matlab toolbox[2] was used for conversion between quaternions and rotation matrices. Due to the missing 3D body joint data from the static trial (MLS) for almost all of the subjects, it was not possible to implement this method in the final program for calculating kinematics for the Kinect MLS. Another problem was that the static measurements, needed to find the rotation between each of the subjects, was not included in the standard output from the MBS. With more time to make changes to the data processing for both systems, a JCS convention for the Kinect data would be possible.

An important point was that the kinematic measures from both systems were as similar as possible with regards to their calculation to get a realistic comparison. We therefore decided to use the estimated joint centers for the hip, knee and ankle that the MBS output already gave us. In the kinematic measurement calculations we defined them according to standard geometry in 3D space. The frontal plane projection angle (FPPA) was also included in the standard output from the MBS (Qualisys), and was compared directly to the same FPPA variable from the MLS (Kinect).

3.5 Kinematic Analysis

The validation process of the MLS (Kinect) was done in form of a kinematic analysis where the values kinematic measurements were compared to the same values from the MBS (Qualisys).

[2]Matlab functions in the field of computer vision by researcher Damien Teney; http://www.montefiore.ulg.ac.be/~dteney/dml.htm

Kinematic output from MBS for two-legged squat

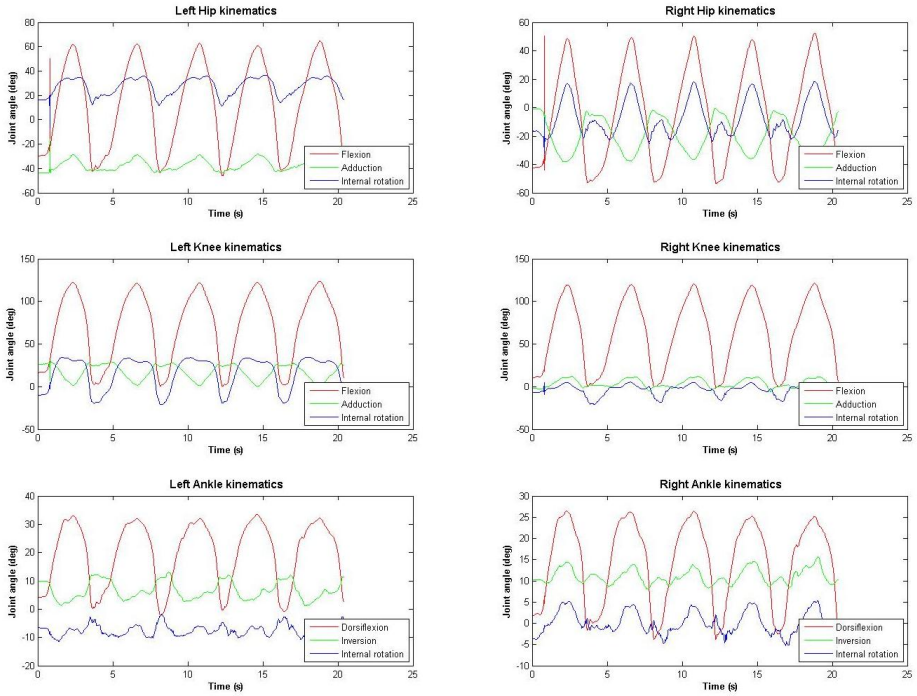

Figure 3.9: Plots of the kinematic output from the MBS (Qualisys) for the two-legged squat task. For each body joint; hip, knee and ankle the three angular measurements flexion, adduction and internal rotation are plotted. These are based on the joint coordinate system (JCS) convention and cannot be directly compared to the MLS (Kinect) output that uses displacement based on strictly 3D coordinate for the body joints.

In section 2.4.2 on page 38 we presented some of the biomechanical measurements used to identify knee instability and injury risk. Since forces were not recorded in the testing, only the kinematic measurements that look at the geometrical motion of the body joints and segments are of interest. The final set of kinematic measurements we chose to use in the analysis of the 3D motion data were relevant for of the application of ACL injury risk screening. It is not given that these measurements are the best for the general evaluation of the pose tracking abilities of the MLS. The contact phase for the vertical drop jump was defined as the phase starting at the initial point of contact(IC) when the subject landed on the floor after dropping from the platform, to the point when the subjects left the floor to perform the jump. It is during the contact phase that the peak knee flexion angle (PF) is extracted.

The tasks and the kinematics measurements of interest are given in the lists below. The frontal plane projection angle (FPPA) measure is illustrated in figure 2.20 on page 39, and in our analysis we used the same definition as presented in the equation in section 2.4.2 on page 39.

The knee-to-ankle separation ratio (KASR) measure was illustrated in a 2D example in figure 2.24 on page 44. In our case the KASR is defined as: the ratio between the magnitude of the 3D vector, from one knee joint center to the other, and the magnitude of the 3D vector between the two ankle joint centers. The KASR is not calculated for the single-leg squat since it requires a two-legged task as mentioned in section 2.4.2.

The knee flexion angle is defined based on the 3D coordinates of the joint centers of the ankle, knee and hip for both systems, and calculated using the standard geometrical method for finding the angle between two vectors in 3D space. The definition of the knee flexion angle in the sagittal plane is illustrated in figure 3.10.

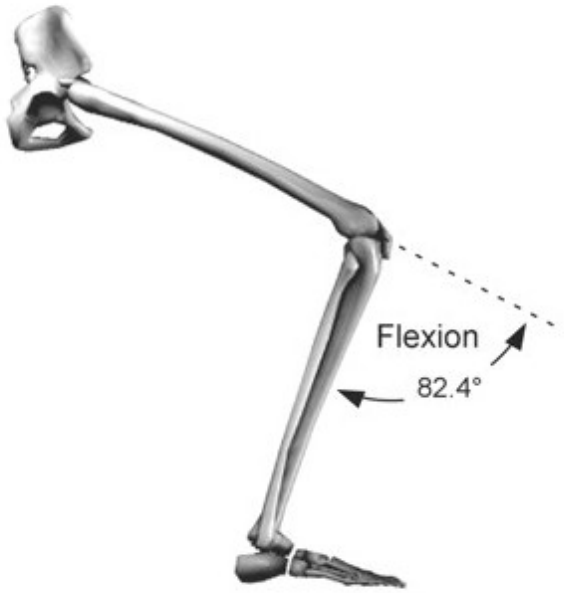

Figure 3.10: Illustration of the knee flexion angle in the sagittal plane.

Tasks for kinematic analysis

1. Vertical drop jump (VDJ)
2. Two-legged squat (front-oriented)
3. Two-legged squat (side-oriented)
4. Single-leg squat (SLS)

Kinematic measurements

a Frontal Plane Projection Angle (FPPA) at IC[3] and PF[4]

b Knee Flexion Angle in 3D

c Knee-to-Ankle Separation Ratio (KASR) at IC and PF

3.5.1 Side Tracking with Microsoft Kinect

As mentioned in the introduction, Kinect v1 is unable to track the skeleton of a body that is oriented sideways. When the user is standing sideways, there are parts of the body that are not visible to the sensor. The skeletal joint positions are not correctly estimated partly because the algorithm is unable to label the required body parts [31]. The depth data from the infrared sensor is used to estimate the joint trajectories in 3D, but this is limited to poses where the user is facing the sensor, which is the normal orientation for a gaming setting. The testing with the Kinect v1 confirmed this, as the skeleton tracking was not close to being accurate for the sideways-recorded two-legged squat task. This motivated an attempt to find a method for tracking joint positions and angles from the side in two-dimensions. The method was partly meant to imitate the visual inspection of two-legged squatting technique, and provide the knee and joint angles which would give a more confident evaluation of the technique in the form of quantified measurements. This work was done before the Kinect v2 sensor was available to use in the testing, but the side-tracking abilities of the Kinect v2 sensor were evaluated after the second testing session.

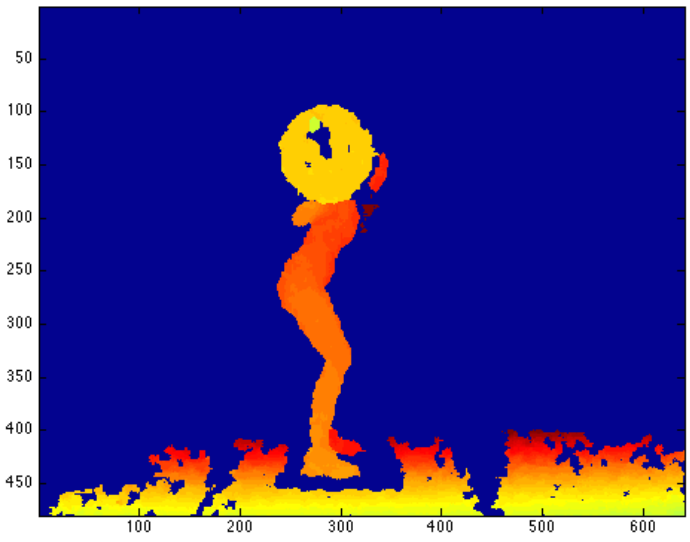

Figure 3.11: Kinect depth image acquired with the IR sensor, exported from Matlab. The color indicates the distance from the sensor.

[3]Initial point of contact; defined by the vertical position of foot marker/joint
[4]Peak Flexion; defined from the knee flexion measurement

A sequence of RBG images and the corresponding depth images from the first session, was the data we had to work with. The subject is performing a series of five continuous squats with bar and weights in the video.

Detecting the Key Points

The first step was to detect the key points defined as the four joint center positions; the hip, shoulder (torso), knee and ankle. To automatically find the location of these points, the body had to be segmented from the background in the image. An example of the depth image obtained from the Kinect video stream is shown in figure 3.11 on the preceding page. The object (person and weights) is already well segmented from the dark background. The blue pixels indicate points which has no depth-value, and the colored pixels in the image have values (in mm) that indicate the distance from the Kinect sensor. An illustration of the Kinect sensor and its components was given in figure 2.2. Since the person was the only object that was of interest, and the weight-plate was about 15-20 cm closer to the sensor than the body, the depth image was perfect for segmentation of the image. To detect the circle-shaped weight-plate, which was the best object to use as a starting point, the Hough transform for circle detection was used.

Hough Transform for Circle Detection

Before applying the Hough transform algorithm to the image in figure 3.11 on the previous page, the noise from the floor around the person was removed. It is usually also a good idea to do some form of pre-processing of the image to smooth out edges, before performing the Hough transform [3]. In this case there was no need for smoothing of edges or morphological operations to fill the gaps.

The Hough transform algorithm that was used [11] is optimized for detecting circles. The function has parameters for the estimated size of the circle, and the minimum number of points required for the circle boundary. It is also possible to specify the expected region of the circle-shape for faster performance, but this was not critical for this application.

The weight-plate was always detected as a circle for the 100 frames that were tested on. Since the algorithm returned a cluster of approximate center points, the center of the cluster was set as the final location of the circle center. The accumulator matrix from the Hough transform result is illustrated in figure 3.13 on the following page, and figure 3.12 on the next page shows the edge-image. The yellow mark indicates the center of the circle that was found.

Segmentation

The original image was segmented based on the depth-value found for the center of the circle, or the nearest point inside the circle with a valid depth-value. To get the best possible segmentation, two thresholds were used to define the depth-range for the side of the body. This created a slice of the depth image going straight through the middle of the body. The visible part of the foot that was furthest away from the camera, was then removed to make the algorithm more accurate. The lower limit of the range of the side of the body was calculated by adding 150 mm to the depth value of the weight-plate, and the upper limit was set by adding 600 mm to the depth value.

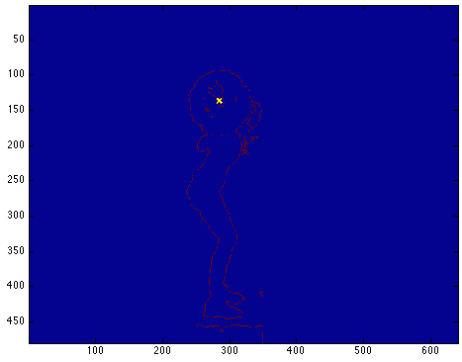

Figure 3.12: Edge image with circle center marked by the yellow cross.

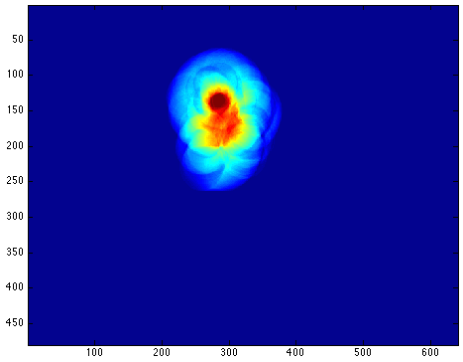

Figure 3.13: The accumulator matrix image from the result of the Hough transform for finding circles.

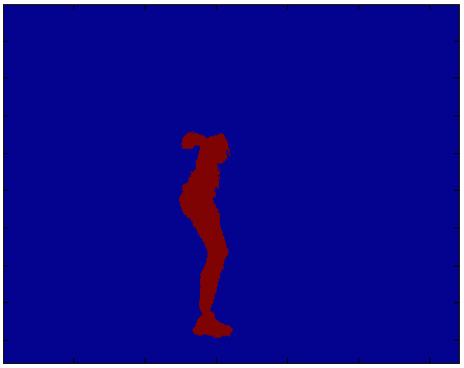

Figure 3.14: The result of segmenting the image with two thresholds on the depth value.

Figure 3.15: The mask obtained from the segmentation was applied to the RGB version of the image for verification. The image illustrates some errors in the segmentation, but was decided accurate enough for our algorithm.

The binary thresholded image is shown in figure 3.14 on the preceding page. The binary mask was applied to the RGB image in figure 3.15 on the previous page to illustrate the segmentation of the real person. There are some areas that are not accurately segmented, for example the hip, but this did not have crucial impact on the performance of the algorithm.

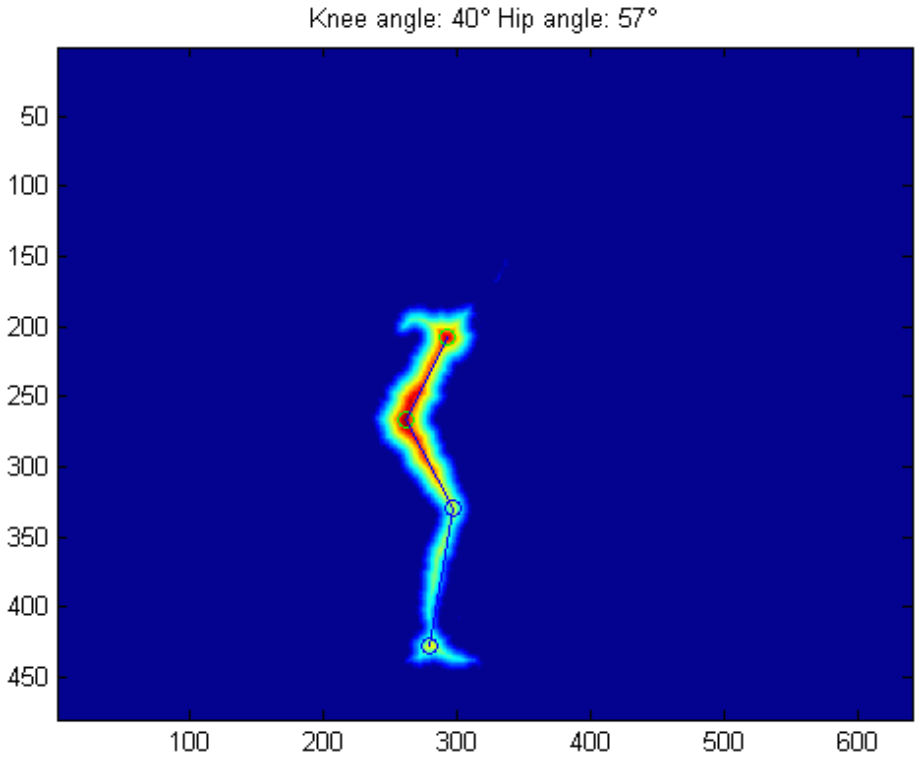

Figure 3.16: The initial joint locations are circled in the distance-image. The hip was defined as the point at the largest distance from the edge of the body. The bone segments (vectors) drawn between the circles.

Distance Transform

The next step after the segmentation was to find the estimated initial position of the joint centers for the shoulder (torso), hip, knee and ankle. The euclidean distance transform for the inverse of the binary image, was computed to create an image where each pixel was assigned a number that indicated the distance between that pixel and the nearest zero-valued pixel. In our segmented image the pixel values corresponded to the distance to the edge of the body.

To find the joint centers, some assumptions were made about their location in the image and the general shape of the body of basketball players. Using the location of the weight-plate as a starting point, the local regions for where to look for the different joints were defined. The hip is obviously at the "thickest" point of the body and corresponds to the pixel with the largest value in the closest pixel map. The torso was assumed to be the maximum distance value

in the region above the hip joint. This was initially thought to be the shoulder joint, but due to occlusion by the weight-plate, we define it as the torso instead.

The knee center was more difficult to find, as there was no single maximum value between the hip center and the ankle center. This problem was solved by dilating the distance-image with a horizontal line as the structure element. The ankle was detected as the maximum distance value in the region below the knee joint. The initial locations of the four joint centers that were detected and the vectors that represent the bone segments, are circled in figure 3.16 on the preceding page.

Angular Kinematics

The anatomical position is zero degrees for the joint angles. The knee and hip angles ($\theta_{knee/hip}$) were calculated using the Law of Cosines shown in equation 3.1 and equation 3.2. a_1, a_2 and b are the two adjacent line segments from the respective joint centers seen in figure 3.17.

$$c^2 = a^2 + b^2 - 2ab\cos\phi_{knee/hip} \tag{3.1}$$

$$\theta_{knee/hip} = 180^o - \phi_{knee/hip} \tag{3.2}$$

Mean Shift Tracking

The joint positions were tracked by using a mean shift tracking algorithm, which is kernel-based and tracks objects in an image based on their appearance [9]. For a sequence of frames, the color histogram and probability density function of the target object (joint) in the current frame is used to create a confidence map for the next image. The pixel values in the map represent the probability that the pixel is from the tracked object. Mean shift is then applied to find the maximum value in the confidence map, that is close to the previous location of the object. As mentioned in the theory chapter, mean shift is used to find modes in the offset joint regression (OJR) algorithm for the upgraded skeletal tracking that comes with the Kinect SDK 2.0.

Implementation

The implementation of the mean shift tracking used a Matlab toolbox, made available by Bernhardt [10], with functions that first creates a mask of a Gaussian-kernel, and then computes the gradients in the horizontal and vertical direction. The code was edited to make it possible to track more than one target at the same time. The kernel profile is used to estimate the probability density function (PDF) for a window around each of the joint positions in the current frame. The estimated PDF for the same location in the next frame is then computed and compared to the current one, by finding similarity value between the two. The mean shift algorithm then uses gradient ascent to find the displacement vector and the new location for the target. A threshold on the similarity value and the allowed number of iterations determines when the algorithm can stop the search.

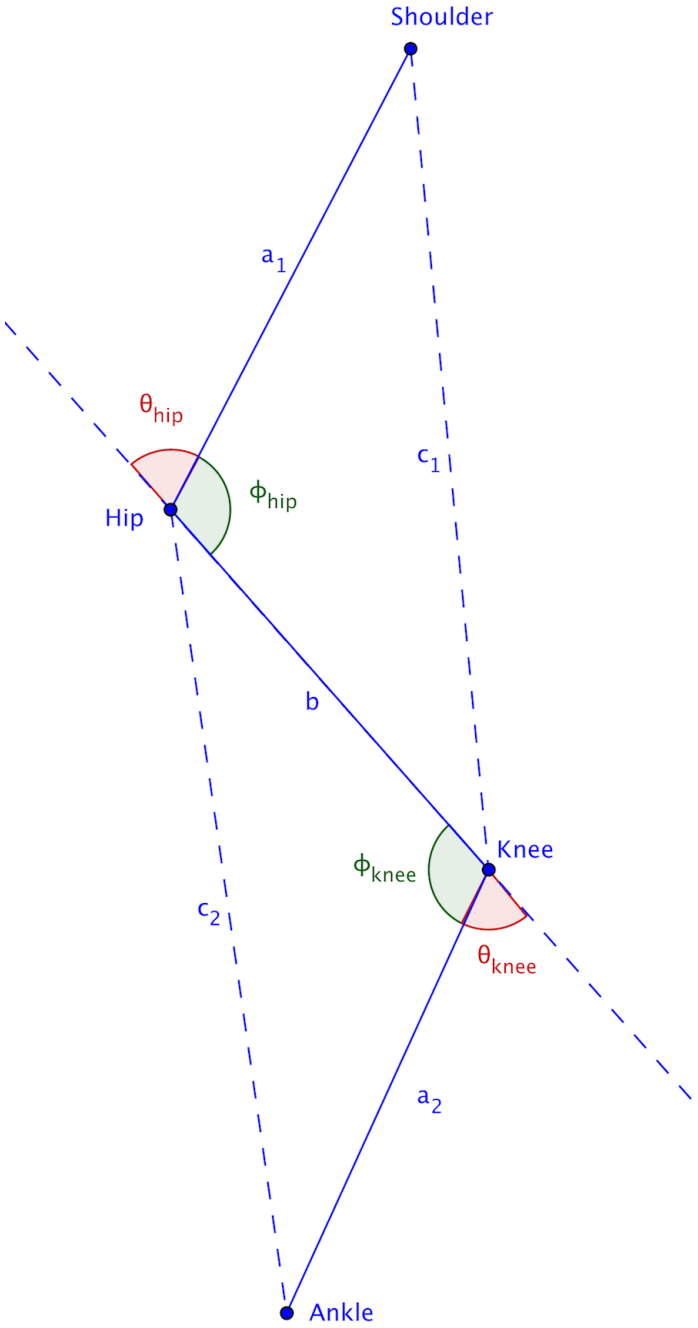

Figure 3.17: Illustration of knee and hip angle calculations in the sagittal plane

3.6 Calibration and Synchronization in Time

One of the secondary goals of this study was to find an efficient method to synchronize the data collected from the two different systems in time. It is not easy to register the exact time that an external event occurs, relative to a video stream. In the case for this experiment, it was a challenge to find the exact time of the start and end of the contact phase for the vertical drop jump task, since there is a significant difference in the frame rate between the two systems. The Kinect sensor has a frame rate of about 30 Hz compared to the infrared cameras of the Qualisys system, which has a stable frame rate of 480 Hz. Different ways of synchronizing the systems in time were discussed in terms of the trade-off between accuracy and the complexity of the testing procedure.

During the preparation for the experiment, it was discovered that the infrared cameras in the Qualisys system were reset and turned off right before the start of the recording. This was visible in the Kinect infrared video stream as a "blinking", which in the best case gives an uncertainty of one frame in the Kinect system. The high framerate of the MBS (Qualisys) makes it difficult to determine exactly when the light is turned on in the captured Kinect frame. For every frame captured by the Kinect sensor, there are $480/30 = 16$ candidates in the markerbased recording for matching the frame that corresponds to the event we wanted to register. Despite the uncertainty due to the unfair difference in frame rates, it was decided that this was accurate enough for the purpose of synchronizing the MLS video stream with the MBS recording in this thesis. A more accurate method for synchronizing the systems in time, with an external trigger that can register events at a higher rate than the Kinect sensor, is presented and discussed later.

The "blink-detection" was implemented in the processing of the raw data from the MLS by calculating the pixel sum in the small area where the IR blink was expected to show for each frame. The two frames with the largest difference in pixel sums were found and the last of these two consecutive frames was identified as the start of the MBS recording. This method was manually verified by looking at the saved infrared images for several of the trials, and finding the first image where "white circles" were visible like it is the image in figure 3.22 on page 67.

3.6.1 Measuring variables independent of time

The challenge of comparing data from two systems with an extreme difference in frame rate and stability encouraged the motivation of investigating the peak values of the knee flexion angle. The peak values can be compared without the need for synchronization in time, and the possible discrepancy between the estimated time of the registered event is not crucial for the motive in this type of setting. This is relevant for the final objective; can the MLS (Kinect) be used in a clinical setting as a simple tool for assessing knee control and stability in athletes. In real-time a Kinect application will be similar to the visual inspection done by physiotherapists to evaluate high knee valgus motion. The Kinect v2 testing session had a different camera setup due to time restrictions and lab availability, so the "blinking" was not visible in the MLS (Kinect v2) recordings.

3.7 Synchronizing in 3D Space

To have a common coordinate system for the data sets, we had to find the translation and rotation between the Kinect coordinate space and the coordinate system defined by the Qualisys system. Both systems had right-handed coordinate frames, and figure 3.18 illustrates the Kinect skeleton space where the positive z-axis is pointing into the direction of view of the sensor. Since

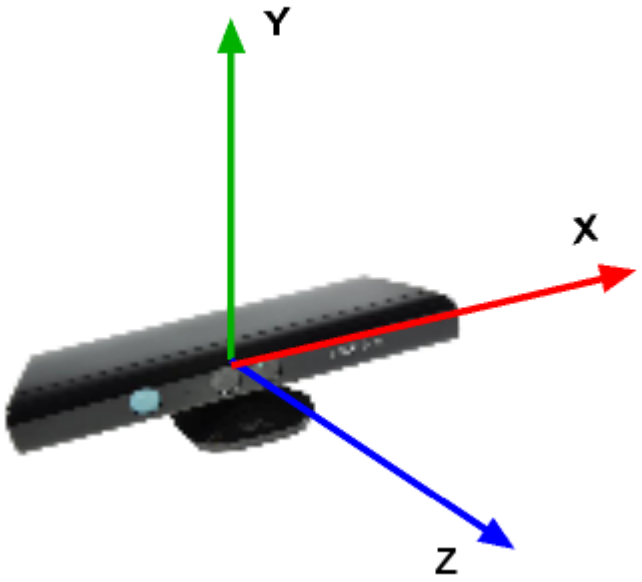

Figure 3.18: Right-handed coordinate system illustrated with the orientation for the Kinect skeleton. A user that is facing the sensor will be looking in the negative z-direction[5].

the Kinect sensor was slightly tilted during the recording, the z-direction was not parallel to the floor. The Kinect coordinate space is oriented relative to the sensor location which means that a user who is facing the sensor, will be looking in the negative z-direction [31]. The skeleton being tracked, is always mirrored, so the user can face their on-screen avatar in the various games making use of the Kinect sensor. The Qualisys coordinate system is right-handed, but the orientation is not the same as for the Kinect skeleton space. A calibration frame with 4 reflective markers was placed in alignment with the force plates on the floor in the motion analysis lab before the initial calibration of the marker based camera system. The photo in figure 3.19 is of the motion analysis lab with the calibration frame in the center, seen from behind the table where the Kinect sensor was placed. The orientation of the coordinate system for the Qualisys recordings were defined in the camera calibration settings of the Qualisys Track

[5]Illustration from http://3dsense.org/

Manager (QTM) application.

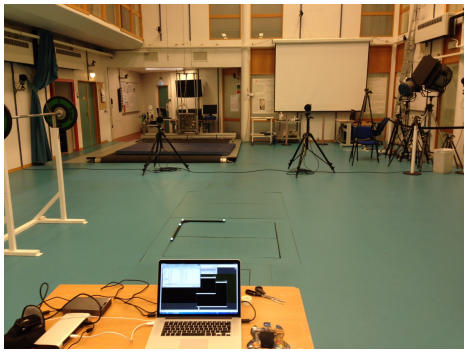

Figure 3.19: Photo of the motion analysis lab setup with calibration frame and the laptop used for recording with the Kinect sensor.

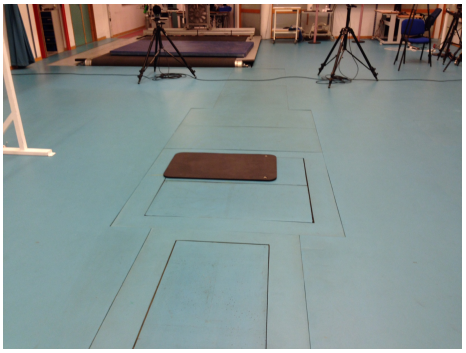

Figure 3.20: Photo of the motion analysis lab after calibration, with landing mat used for better depth estimation during the tasks.

The positive x-axis was represented by the long arm of the calibration frame seen in figure 3.19, that had a middle marker, and an end marker, and the short arm with the single end marker corresponded to the positive y-axis. The positive z-axis was pointing upwards with the origin in the marker in the corner of the calibration frame.

3.7.1 Aligning the 3D Models

In figure 3.21 on the following page the skeletons from the two 3D models we wish to compare, are shown on top of each other to give a visual impression of what we wanted to achieve by finding the transformation between the MBS and the MLS coordinate systems.

Since the marker positions are used to draw the skeleton from Qualisys, they do not correspond to the joint positions that are drawn for the Kinect skeleton.

A soft landing mat was placed on top of the force plates in the first screening session, as this improved the capture of the depth data with the Kinect v1 sensor. The reason we got better depth data acquisition with the mat is probably related to the light-sensitivity of the Kinect v1 sensor since the light in the room was reflected off of the smooth floor in the lab. Adding the mat was possible since it was not necessary to measure ground reaction forces in this experiment. The Kinect v2 showed no visible problem with capturing depth so it was not necessary to use the mat in the second screening session, The setup for the lab after calibration is seen in the photo in figure 3.20.

3.7.2 Transformation using marker based calibration frame

The transformation between the two coordinate spaces consisting of translation and rotation had to be estimated in order to align the skeletons. The strategy for finding this rigid transformation included using an infrared image of the calibration frame, captured by the image stream from the Kinect sensor.

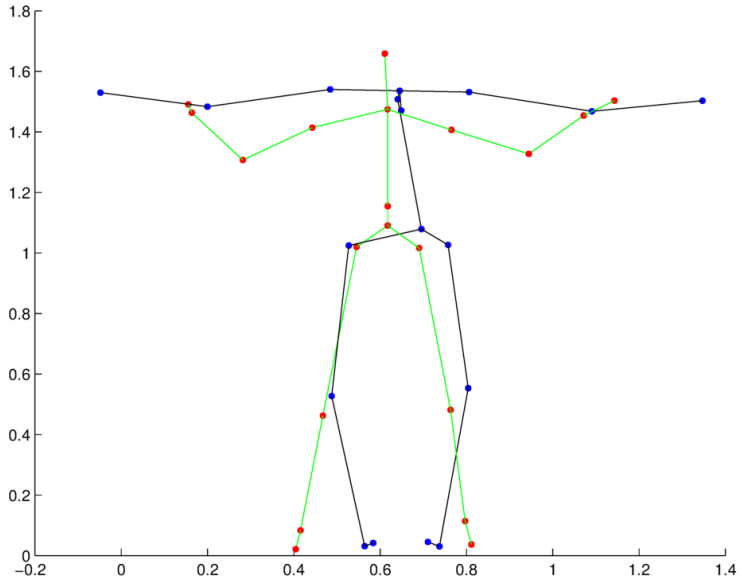

Figure 3.21: Result of manual skeleton alignment with the 20 joints of the Kinect skeleton, illustrated with red dots and bone segments in green. The black skeleton is the skeleton created from the Qualisys marker positions. The center of the hip is here replaced by the posterior left hip marker position.

One of the infrared images of the calibration frame that was captured, is shown in figure 3.22 on the following page, where the four infrared markers on the frame are clearly visible.

The procedure for finding the translation involved using RANSAC[6] to find the equation for the floor-plane, using the points that lay within the triangle defined by the markers in the infrared image. This technique did not give a transformation that was accurate enough for further implementation.

3.7.3 Kinect Calibration Toolbox

Another method we experimented with to find the transformation between the data, made use of the calibration values for a Kinect v1 sensor, identical to the one that was used in this master project. In 2012 Herrera released a Matlab calibration toolbox for the Kinect, that implemented a algorithm calibrating a depth camera and two color cameras at the same time [19]. Even though each Kinect sensor can vary in its camera intrinsics, the documentation for the toolbox stated that "calibration results should not vary very much between Kinects"[19], so the predefined calibration values from the toolbox sample was used in the attempt to estimate the rigid transformation. Using these calibration values are also justified by the fact that the official Kinect SDK automatically calibrates the sensor. When the transformation was applied to the 3D data, the skeletons still did not align properly, so we chose to discard this and focus on other more important objectives of the thesis.

[6]RANSAC; iterative method used to estimate parameters of a mathematical model

Figure 3.22: Infrared image of the MBS calibration frame captured by the Kinect v1

3.8 Statistical Measurements and Data Analysis

SPSS Statistics (version 22) was used for the statistical analysis of the kinematic measurements. The marker based method of performing motion analysis for screening for ACL injury risk has been established as the "gold standard" in this field of research and testing. In the data analysis we defined the results from the MBS (Qualisys) as the truth, and compared difference in measurements from this system to those of MLS using standard statistical methods.

The kinematic measures that we evaluated and analyzed statistically, were based on a basic geometrical representation for angular and linear motion in 3D. In the kinematic measures the knee flexion angle (PF), frontal plane projection angle (FPPA) and knee-to-ankle separation ratio (KASR, defined in the theory chapter p. 40) were calculated from the lower body joint center positions. The statistical variables were extracted from these measurements for each leg; from all trials for all tasks for each subject. The vertical drop jumps (VDJ) were recorded as separate trials, but the trials for the other tasks were performed in a continuous motion, and had to be separated. The contact phase for the VDJ was defined earlier as the period in time from initial contact (IC) with the floor to the time the feet leaves the floor. The contact phase defines the time axis in the plots of the variables for the VDJ. For the other tasks, the IC was defined as the start of the recording since the feet never left the ground.

Comparing Means

For all tasks we compared the mean values (averaged over the number of trials) for each variable using the paired samples t-test. This test determines whether there is a significant difference between the measured values assuming indepen-

dent and identically normally distributed paired differences. The paired samples t-test was applied to the mean values (average over all trials for each subject), of all kinematic variables, at the time of peak flexion (PF).
In addition to see if there is a significant difference between the mean of the variables within each of the systems, we also compared the mean values for the left leg to those of the right leg. This was to see how large the difference between the left and right leg was in the markerless system (Kinect) compared to the marker based system (Qualisys). This comparison was only done for the two tasks VDJ and the two-legged squat, where the data for both leg was from the same recording, and we expected no significant difference from one leg to another. The applicable kinematic measures for this inter-system comparison were; peak flexion (PF) and frontal plane projection angle (FPPA). The knee-to-ankle separation ratio (KASR) was not included since it as mentioned earlier, needs joint positions from both legs to produce one measurement value.

Rank Correlation

We needed a way of comparing the results from the kinematic measurements that would relate it to knee motion analysis. In other words we wanted to see if the MLS would give subjects with high knee motion (large angular displacement) the same rating as the MBS. The statistical measure we decided to use for this was Spearman's rank correlation coefficient (SRCC), which outputs a value in the $[-1, 1]$ range. The SRCC (ρ) is a non-parametric measure that evaluate the statistical dependence between the variables, to see if the relationship between them can be described using a non-linear monotonic function [42]. The difference between the SRCC and the Pearson correlation coefficient is that SRCC ranks the variables according to their position in ascending order. This way we avoid Pearson's assumption of linear correlation which is not the case for our variables as indicated in the scatter plot figure 3.23 on the next page. The mean peak flexion value for the left leg for each subject for the vertical drop jump is shown in the plot. The SRCC looks to see if the high data values for the MBS are also ranked as high values in the MLS, and is not as sensitive to outliers as a parametric method such as the Pearson coefficient [42].

Correlation coefficients of 0.81 to 1.00 were interpreted as being very good, 0.61 to 0.80 as good, 0.41 to 0.60 as moderate, 0.21 to 0.40 as fair, and less than 0.20 as poor [4].

3.8.1 Evaluating the Reliability of MLS

A good way to measure the reliability of a MLS in the type of experiment setting we had, would be to have a re-test session with the same subjects a few weeks or months after the first session. We had to stick to evaluating the reliability of the Kinect v1 sensor at a inter-trial level and compare it to that of the MBS (Qualisys). We used the intra-class correlation coefficient (ICC) to describe the consistency of measurements made by the systems between the individual trials. We used the two-way random (absolute agreeement) type of ICC.

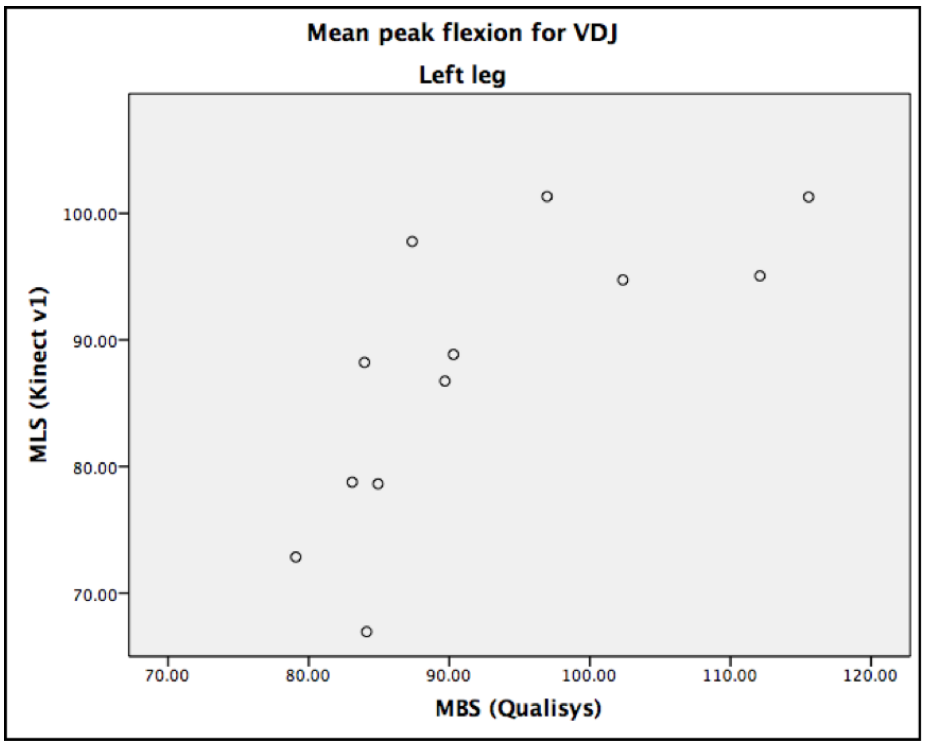

Figure 3.23: Scatter plot of two paired peak flexion (PF) variables from the MLS (Kinect v1) and the MBS (Qualisys). The plotted data points indicates a non-linear correlation between the variables (exported from SPSS).

Chapter 4

Results

The experiments described in the material and method chapter were the basis of a quantitative kinematic analysis for validation of the markerless motion capture system (MLS) Kinect for evaluating sports movement. The main objective was to see if MLS is accurate enough to evaluate knee motion of female athletes during three front-oriented sports movements; vertical drop jump (VDJ), two-legged squat and single-leg squat (SLS). Results from the joint tracking algorithm developed for evaluation of hip and knee motion during side-oriented squats are included. The accuracy of the Kinect was assessed based on the results from a statistical analysis comparing the mean values of the kinematic variables.

4.1 Data Sets and Descriptive Statistics

After the data processing for all screening sessions we discovered errors in the data from both systems. This caused a small reduction in the total size of some of the data sets that were input for the statistical analysis. The results will be presented in this chapter with relevant tables and plots, and the errors discovered in the data sets will be discussed later. The raw data from MLS (Kinect v1 and v2) was processed in Matlab and the joint positions were defined as variables for each trial for each subject. The corresponding MBS data for each trial was loaded in, and kinematic variables was calculated for each trial for both systems using the same functions. We ended up with a set of discrete variables as input in the statistical analysis.

4.1.1 Comparing Kinect v1 and Qualisys

The kinematic variables from session 1 with descriptive statistics are listed in table 4.1 on the following page . The mean and standard deviation (S.D.) of markerless (MLS) and marker based (MBS) kinematics, during three tasks at two time points were calculated. The MLS values marked in red are significantly different from the MBS data (two-tailed paired samples t-test, $p < 0.05$). The tasks and kinematic measures were presented in chapter two with the mathematical details of the calculations of PF and FPPA angles and KASR.

Table 4.1: *Session 1*: Mean and standard deviation (S.D.) of markerless (MLS) and marker based (MBS) kinematics for 12 (squat; 10) subjects, during three tasks at two time points; vertical drop jump (VDJ), two-legged squat (squat) and single-leg squat (SLS); peak knee flexion (PF), frontal plane projection angle (FPPA) and knee-to-ankle separation ratio (KASR); at initial contact (IC) and peak flexion (PF). The MLS values marked in red are significantly different from the MBS data (two-tailed paired samples t-test, $p < 0.05$).

Descriptive Statistics for Session 1; Mean (± S.D.)						
Variables			MLS (Kinect v1)		MBS (Qualisys)	
Task	Kinematics	Time point	L	R	L	R
VDJ	PF		87.6 (11.2)	80.3 (10.4)	92.5 (11.8)	88.4 (8.8)
	FPPA	at IC	5.1 (3.9)	3.1 (3.9)	4.4 (3.9)	2.2 (3.9)
		at PF	2.4 (8.6)	11.0 (9.5)	0.2 (8.6)	7.2 (15.1)
	KASR	at IC	0.82 (0.06)		0.86 (0.05)	
		at PF	1.02 (0.08)		0.86 (0.13)	
Squat	PF		92.2 (7.4)	96.2 (8.8)	107.0 (7.2)	105.6 (5.1)
	FPPA	at PF	46.6 (11.5)	55.2 (13.3)	55.7 (16.2)	53.7 (17.9)
	KASR	at PF	1.06 (0.11)		1.13 (0.13)	
SLS	PF		82.6 (13.7)	89.7 (13.5)	86.1 (4.6)	110.1 (7.3)
	FPPA	at PF	-7.0 (15.0)	20.9 (16.6)	-7.8 (9.7)	37.6 (25.2)

4.1.2 Comparing Kinect v2 and Qualisys

Session 2 was only a pilot-testing with three subjects, and the results from the statistical analysis will only give us an indication of the Kinect v2 sensor's performance on a larger data set. The MLS and MBS data from session 2 were not synchronized to have the same starting point in time which influences the plotted kinematic measures. The time axis in the plots is defined based on the point in time of initial contact (IC) and the end of the contact phase for the VDJ. Descriptive statistics for the discrete variables in session 2 are listed in table 4.2

Table 4.2: *Session 2*: Mean and standard deviation (S.D.) for 3 subjects. Markerless (MLS) and marker based (MBS) kinematic variables; peak knee flexion (PF), frontal plane projection angle (FPPA) and knee-to-ankle separation ratio (KASR) were extracted from two tasks; vertical drop jump (VDJ) and two-legged squat (squat) at initial contact (IC) and peak flexion (PF). The MLS values marked in red are significantly different from the MBS data (two-tailed paired samples t-test, $p < 0.05$).

Descriptive Statistics for Session 2; Mean (± S.D.)						
Variables			MLS (Kinect v2)		MBS (Qualisys)	
Task	Kinematics	Time point	L	R	L	R
VDJ	PF		88.6 (8.9)	94.5 (9.1)	80.3 (7.2)	80.3 (8.8)
	FPPA	at IC	9.5 (6.4)	6.9 (2.6)	29.4 (8.9)	33.3 (0.8)
		at PF	20.6 (10.7)	34.8 (21.7)	80.3 (7.2)	80.2 (8.9)
	KASR	at IC	0.91 (0.02)		0.86 (0.03)	
		at PF	0.78 (0.14)		0.88 (0.14)	
Squat	PF		109.0 (8.6)	108.1 (8.0)	110.9 (15.1)	109.7 (12.0)
	FPPA	at PF	87.9 (16.6)	88.5 (15.5)	114.3 (19.1)	110.7 (13.4)
	KASR	at PF	1.06 (0.06)		1.10 (0.06)	

4.2 Accuracy and Tracking Ability of Kinect MLS

The final output of the statistical analysis of the data from session 1 is listed in table 4.6 on page 87. The mean values of all variables were compared and is presented as p-values in the table. Green p-values indicate that no significant difference was found between the MLS and the MBS (two-tailed paired samples t-test, $p < 0.05$). The values marked with red are the cases where the mean variables differed significantly.

Spearman's Rank Correlation Coefficient (SRCC)

Spearman's rank correlation coefficient (SRCC) is a non-parametric measure that we used to evaluate the statistical dependence between the variables from the MLS and the MBS. The SRCC results are given as p-values in table 4.6 on page 87, where the green cells mark values from 0.600 and up. The $*/**$ indicates correlation that is significant at the 0.05/0.01 level (2-tailed). This SRCC results are relevant for the evaluation of ability of the MLS to rank subjects with high knee motion with the same accuracy of the MBS. Variable with very good correlation are marked in dark green (SRCC: $0.81-1.0$). Light green correlation values are interpreted as good (SRCC: $0.61-0.80$), and yellow equals moderate correlation (SRCC: $0.41-0.60$). 0.41 to 0.60 as moderate. The red values indicate lower significance and poor (or fair) correlation (SRCC < 0.40).

Knee-to-ankle Separation Ratio (KASR)

The KASR is only applicable for two-legged tasks, and was outputted as a separate plot for each trial. A positive KASR indicates a varus knee angle, "bow-legged", where the magnitude of the vector between the knee joints is

larger than between the ankles. A negative ratio represents a valgus knee angle, "knock-kneed", in the case where the knees collapse together and are closer to each other than the ankle joints. The vectors in the KASR calculation are illustrated in chapter two in figure 2.24 on page 44. We evaluated the KASR based on the absolute value, since a large magnitude ratio indicates high knee motion in both the varus and valgus cases. The varus and valgus knee stances are illustrated in figure 4.1.

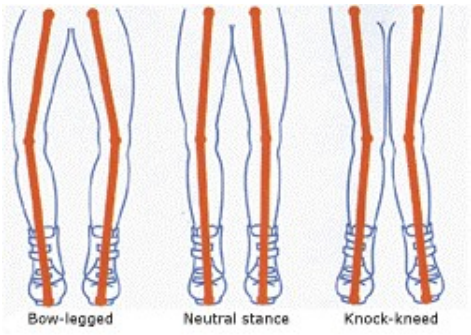

Figure 4.1: The varus knee angle, "bow-legged" is shown to the left and the valgus knee angle, "knock-kneed", is to the right of the neutral stance[1]

4.2.1 Inter-trial reliability

The reliability of MLS is assessed at an inter-trial level in this project. Since the repeated trials are from the same session, the results does not say much about the repeatability of the Kinect sensor. The vertical drop jumps were recorded as separate trials, and for the other tasks the five trials were performed in a continuous motion and had to be separated. The degree of agreement between the measured values for each trial was evaluated using the intra-class correlation coefficient (ICC).

The results from testing the inter-trial reliability of the MLS are given by the ICC values in table 4.6 on page 87. The color-coding of the values are defined as for Spearman's correlation coefficient; from dark green to red (very good, good, moderate and poor/fair).

Session 2 with Kinect v2

The results from the statistical analysis for session 2 in listed in table 4.3 on the following page. KASR at IC and PF for the VDJ showed consequently no significant difference for the mean comparison.

For session 2 the intra-class correlation coefficient (ICC) was calculated for the mean peak flexion variable for vertical drop jump. The ICC for the four trials of VDJ, are presented in table 4.4 on the next page for MLS (Kinect v2) and MBS (Qualisys).

[1]Illustration from: http://buddingbodyworker.blogspot.no/

Table 4.3: *Session 2*: Statistical Analysis of markerless (MLS) and marker based (MBS) kinematics, during three tasks at two time points; vertical drop jump (VDJ), two-legged squat (squat) and single-leg squat (SLS); peak knee flexion (PF), frontal plane projection angle (FPPA) and knee-to-ankle separation ratio (KASR); at initial contact (IC) and peak flexion (PF). The values marked in red indicate a significant difference between the MLS and MBS for the compared means, and green means no significant difference was found (two-tailed paired samples t-test, $p < 0.05$). Spearman's Rank Correlation Coefficient (SRCC) values cannot be interpreted due to small sample size.

Statistical Analysis Session 2; MLS (Kinect v2) vs MBS (Qualisys)								
Comparing Means; Paired Samples T-test							SRCC	
Variables			MLS vs MBS		MLS	MBS	MLS vs MBS	
Task	Kinematics	Time point	L	R	L - R	L - R	L	R
VDJ	PF		0.0140	0.0805	0.2227	0.9959	1.0	0.5
	FPPA	at IC	0.0131	0.0031	0.5224	0.4899	0.5	0.5
		at PF	0.0023	0.1032	0.4130	0.9383	1.0	-0.5
	KASR	at IC	0.2029				-0.5	
		at PF	0.5530				-0.5	
Squat	PF		0.6974	0.6883	0.3555	0.5923	0.5	0.5
	FPPA	at PF	0.0257	0.0483	0.8607	0.4086	0.5	0.5
	KASR	at PF	0.0811				0.3	

ICC for Peak Flexion during VDJ			
MLS (Kinect v2)		MBS (Qualisys)	
L	R	L	R
0.472	0.028	0.826	0.859

Table 4.4: *Session 2*: Table with intra-class correlation coefficients(ICC) for MLS (Kinect v2) and MBS (Qualisys) for peak flexion for vertical drop jumps.

ICC for Peak Flexion during Squat			
MLS (Kinect v2)		MBS (Qualisys)	
L	R	L	R
0.959	0.945	0.990	0.980

Table 4.5: *Session 2*: Table with intra-class correlation coefficients(ICC) for MLS (Kinect v2) and MBS (Qualisys) for peak flexion for front-oriented squats.

4.2.2 Knee Motion Tracking - Vertical Drop Jump

The results for the vertical drop jump (VDJ) are presented in form of discrete variables in table 4.3 on the preceding page for session 1 and table 4.3 on the previous page for session 2. The vertical drop jump (VDJ) measurements had an additional statistic variable at the time of initial contact (IC).

Session 1 - Kinect v1 MLS

The only VDJ variable that was significantly different from the MBS, was mean peak flexion (PF) for the right leg (two-tailed paired samples t-test, $p < 0.05$). There was a significant difference between the mean PF values for the left leg compared to the right leg in the MLS (Kinect v1). This was also the case for the frontal plane projection angle (FPPA) at the time of peak flexion (PF). The FPPA comparison of means between the systems had significantly higher p-values for the right leg than the left leg. Spearman's rank correlation coefficient (SRCC) for VDJ indicated significant correlation between variables at the 0.05 level for the left leg (PF and FPPA). FPPA at PF for the right leg also had a high SRCC value.

The knee flexion angle and FPPA measures are plotted for one subject in session 1 in figure 4.2, and the KASR in figure 4.3 on the next page.

Session 1 - Knee flexion and FPPA (VDJ)

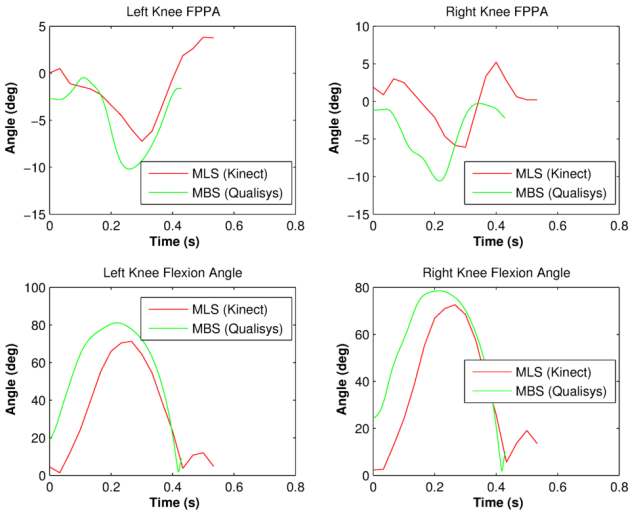

Figure 4.2: *Session 1*: The FPPA and knee flexion angular displacement during vertical drop jump (VDJ) for one subject. Both variables show clear signs of correlation between MLS and MBS. The data is only synchronized at the time of initial contact (IC) defined individually for each system.

Session 1 - Knee-to-ankle separation ratio (VDJ)

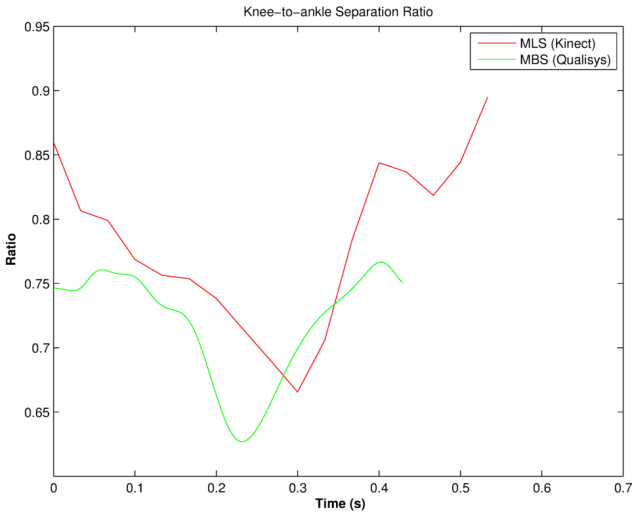

Figure 4.3: *Session 1*: Knee-to-ankle separation ratio (KASR) during vertical drop jump (VDJ) for one subject. The curves are similar in shape, but there is a difference in value of the ratio at the time of peak flexion. The data is only synchronized at the time of initial contact (IC) defined individually for each system.

Session 2 - Kinect v2 MLS

Almost all mean values for Kinect v2 MLS were significant different from MBS. The exception was the KASR measure (at IC and PF) that both had high p-values. The SRCC was -0.5 for both KASR variables which indicates a decreasing monotonic relationship. The KASR for one subject in session 2 during the contact phase of VDJ is plotted in figure 4.5 on the following page, and can in the best case confirm significant correlation although the data is not synchronized for optimal comparison.

The plotted knee flexion-curve in figure 4.4 on the next page during VDJ looks highly correlated as well, which is promising in regards to accuracy for this measure.

Session 2 - Knee flexion and FPPA (VDJ)

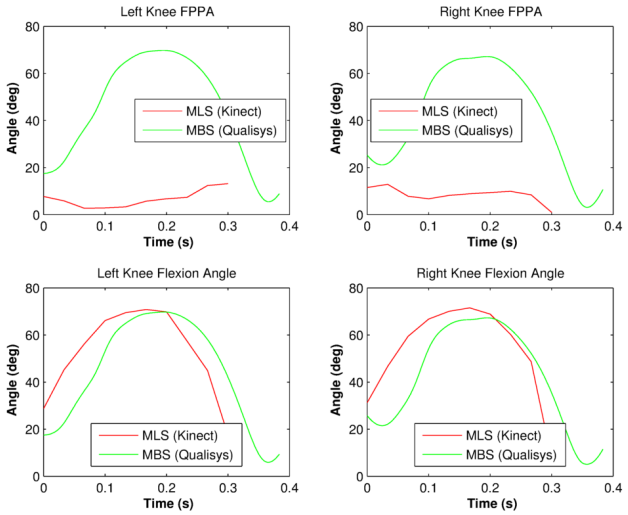

Figure 4.4: *Session 2*: The FPPA and knee flexion angular displacement during vertical drop jump (VDJ) for one subject. The FPPA curve for MLS is flat and did not give any valid results for any of the subjects in session 2. The knee flexion for both legs show signs of a correlation between the MLS and MBS. The data is only synchronized at the time of initial contact (IC) defined individually for each system.

Session 2 - Knee-to-ankle separation ratio (VDJ)

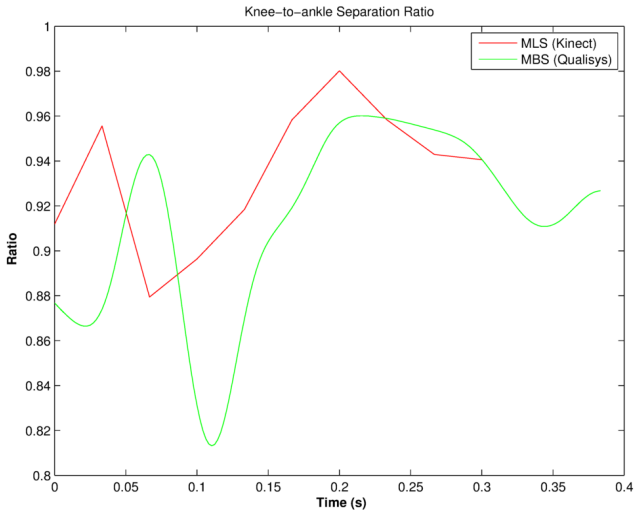

Figure 4.5: *Session 2*:Knee-to-ankle separation ratio (KASR) during vertical drop jump for one subject. The data is only synchronized at the time of initial contact (IC) defined individually for each system.

4.2.3 Knee Motion Tracking - Front-oriented Squat

Session 1 - Kinect v1 MLS

The mean of PF for the front-oriented squat showed a significant difference for both legs, according to extremely low p-values from the result of the t-test. The KASR rank correlation was significant at the 0.01 level for the squat at PF according to the SRCC value (0.927). The other squat correlation values were also in the upper interval of the SRCC measurement. The FPPA and knee flexion angular displacement during five consecutive two-legged squats (front-oriented) are plotted in figure 4.6, and the KASR for the same task is plotted in figure 4.7 on the next page. The plotted variables are only synchronized in time based on the start of the MBS recording.

Session 1 - Knee flexion and FPPA (squat)

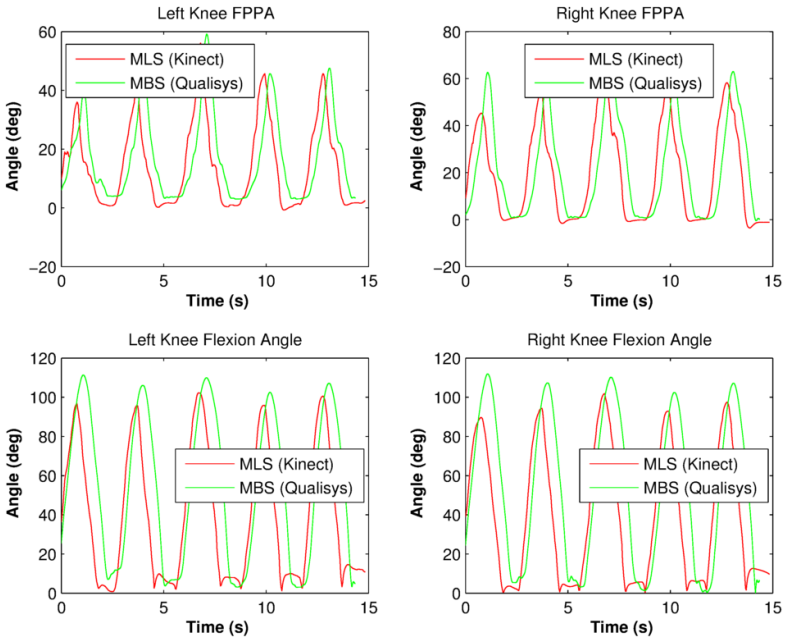

Figure 4.6: *Session 1*:The FPPA and knee flexion angular displacement during five consecutive two-legged squats (front-oriented) for one subject. The MLS has smaller peak values than the MBS for both variables for both legs. The data is only synchronized in time at the starting point.

Session 1 - Knee-to-ankle separation ratio (squat)

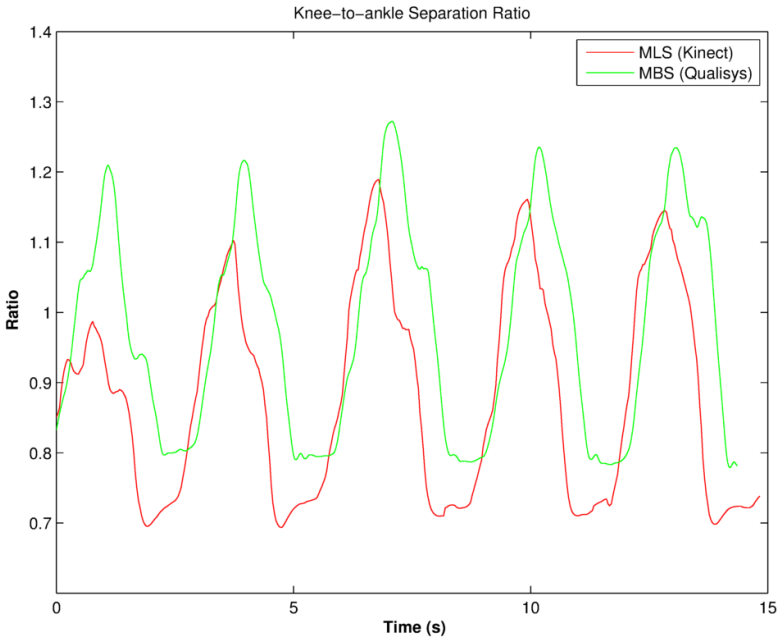

Figure 4.7: *Session 1*: Knee-to-ankle separation ratio (KASR) during five consecutive squats (front-oriented) for one subject. The MLS has smaller peak values than the MBS, but follows the pattern of the MBS for all trials. The data is only synchronized in time at the starting point.

Session 2 - Kinect v2 MLS

The front-oriented squat had very good results across the board for the peak flexion (PF) variable. The strong relationship between the Kinect v2 and the MBS PF variable is seen in figure 4.8 on the following page. The KASR plotted in figure 4.8 on the next page was also promising in terms of shape, but Kinect v2 underestimated all peak values compared to MBS.

Session 2 - Knee flexion and FPPA (squat)

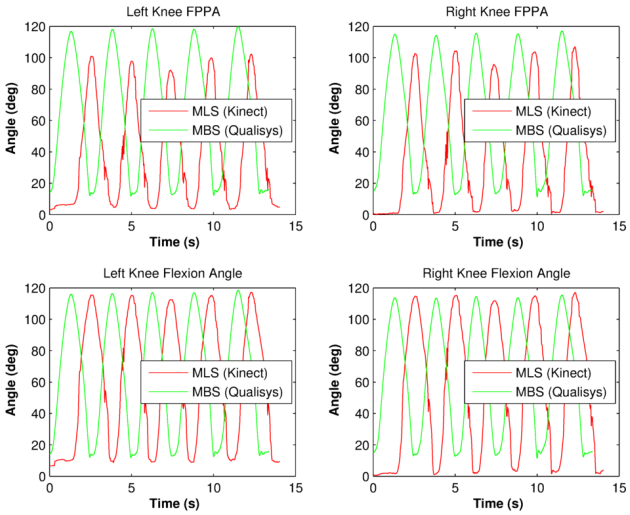

Figure 4.8: *Session 2*: The FPPA and knee flexion angular displacement during five consecutive two-legged squats (front-oriented) for one subject. MLS (Kinect v2) have smaller peak FPPA values than MBS, but MLS peak flexion seem to be about the same as MBS for both legs. The data is not synchronized in time, so there is a large offset due to an earlier start of recording with for MLS.

Session 2 - Knee-to-ankle separation ratio (squat)

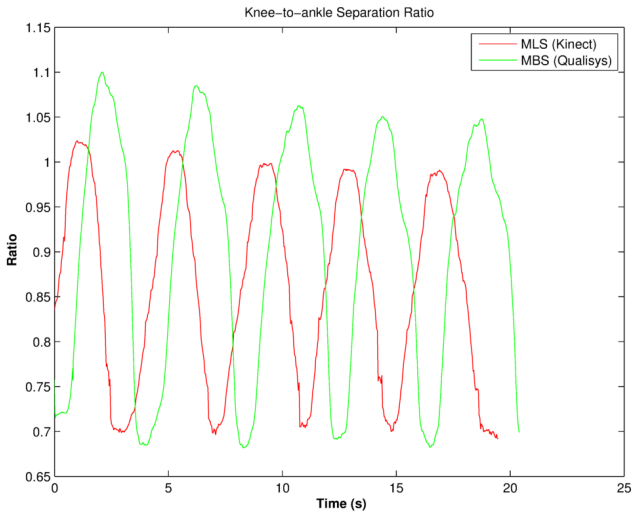

Figure 4.9: *Session 2*: Knee-to-ankle separation ratio (KASR) during five consecutive squats (front-oriented) for one subject. The MLS has smaller peak values than the MBS, but follows the pattern of the MBS for all trials. The data is not synchronized in time, so there is a large offset due to different starting points in time for each system.

4.2.4 Knee Motion Tracking - Single-leg Squat

Results for the single-leg squat was not obtained from session 2 due to error in the processing. Mean values of SLS for MBS seen in table 4.1 on page 72 were not as expected, and comparing them with MLS mean values gave a poor result overall seen in table 4.6 on page 87. The results from the paired samples t-test for single-leg squat (SLS), showed a significant difference in the compared means of PF for the right leg. There was no significant difference between the markerless (MLS) and the marker based system (MBS) for the left leg in for PF, and both legs for FPPA at PF. The p-values for the left leg in this task (PF and FPPA) indicated rejection of the null-hypothesis, and no significant difference between the MLS and MBS. The SRCC results for SLS were all below 0.400.

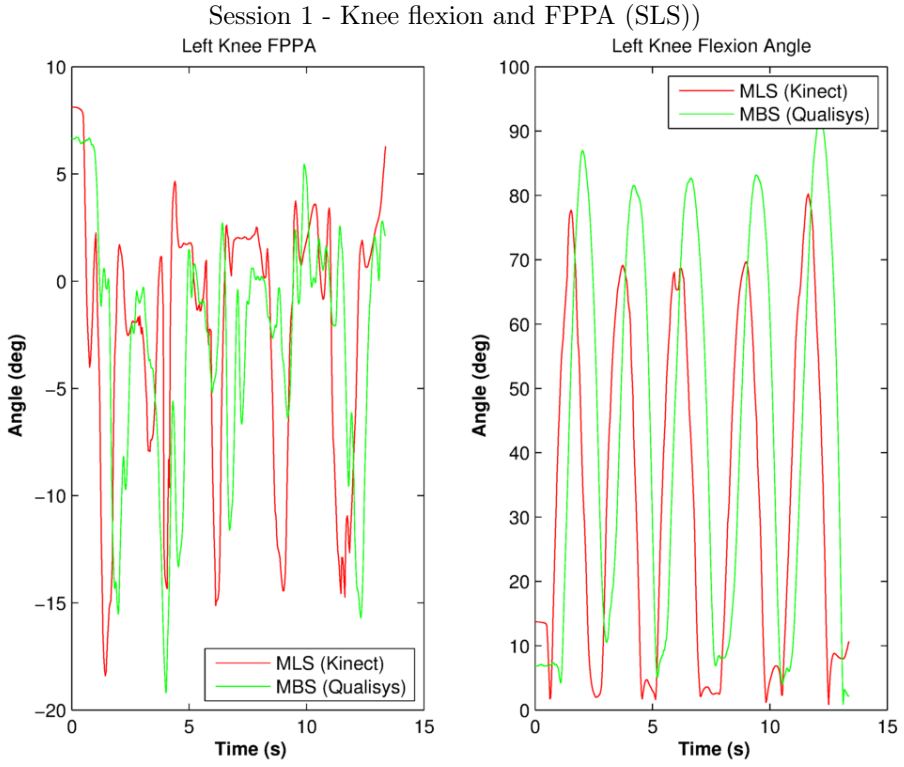

Figure 4.10: *Session 1*: Comparison of MLS and MBS Knee flexion and frontal plane projection angles (FPPA) during five consecutive single-leg squat (SLS) trials. The MLS have consistently smaller peak knee flexion angles than the MBS. The FPPA measure show similar peak values for few of the trials.

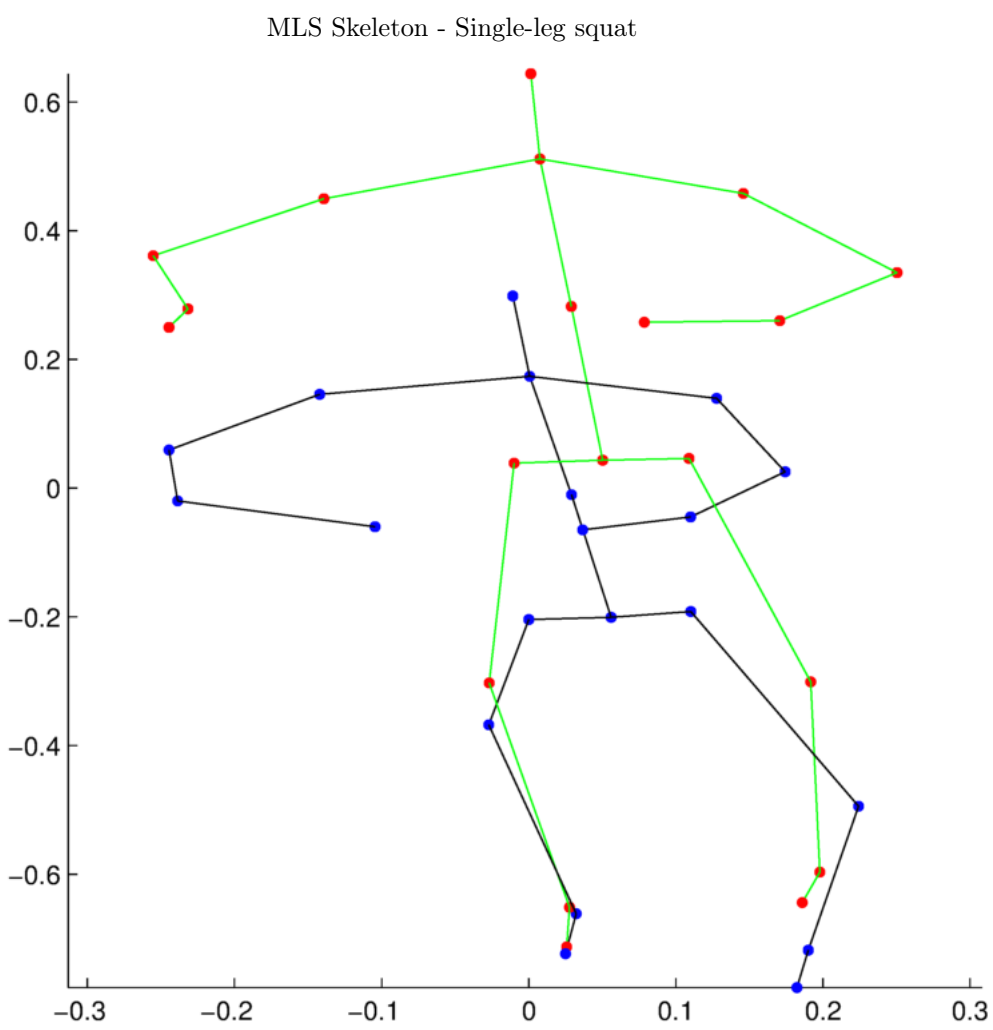

Figure 4.11: MLS skeletons in the frontal plane; the green skeleton is at the beginning of the task, and the black is at peak flexion (PF) during a single-leg squat (SLS).

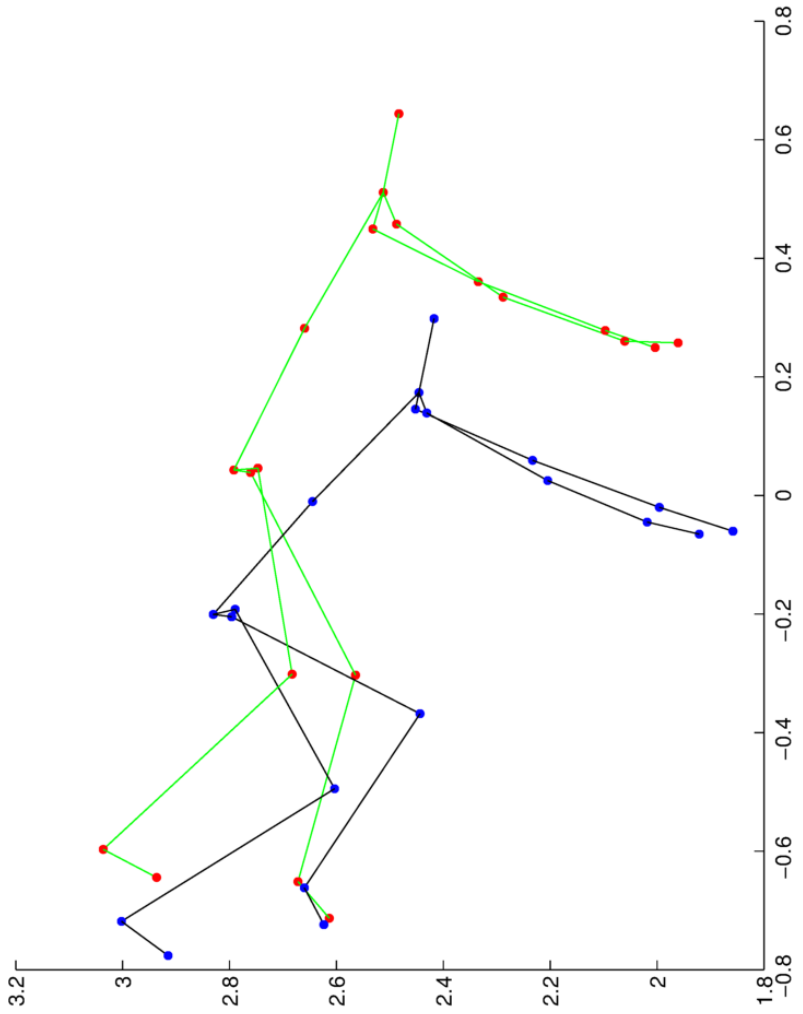

Figure 4.12: MLS skeletons in the sagittal plane; the green skeleton is at the beginning of the task, and the black is at peak flexion (PF) during a single-leg squat (SLS).

4.2.5 Motion Tracking - Side-oriented Squat

Our side tracking algorithm that used image analysis techniques to estimate joint angles, produced the results seen in figure 4.14 on the following page. The knee and hip angles are plotted before and after using a smoothing filter (1st degree polynomial model) on the data. The algorithm tracked the hip and ankle positions fairly well throughout the video sequence of about 400 frames. The tracked knee joint location was not consistent, and gradually moved away from the correct position shifting up and down more than expected.

The filtered x-and y-positions of the tracked joints in a series of 5 squats, are plotted in figure 4.15 on the next page. The tracking fails to stay with the correct position of the shoulder joint (torso) after 3 reps when the player is on her way back up from the squatting position. The knee joint has too much variation in the vertical direction, and the additional dip in the y-plot indicates error in the tracking. An image of the athlete in the deepest position for the squat, is shown in figure 4.13, and from visual inspection the joint positions seem to be correctly estimated. The peak flexion angles for the knee and hip are found at the deepest squat position. A video of the result of the joint and angle tracking can be found at the UiO student website[2].

Side tracking of hip and knee angles (Kinect v1)

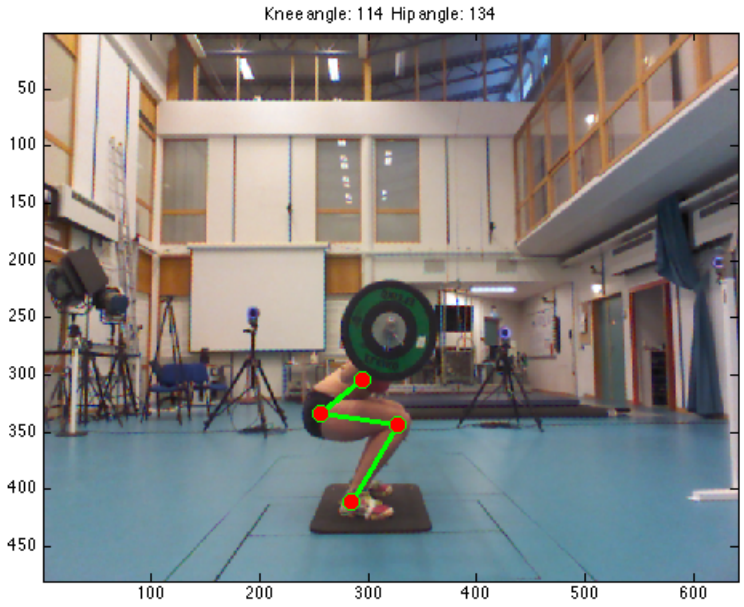

Figure 4.13: Joint position tracking at the deepest squat position

[2]http://folk.uio.no/tinasmo/

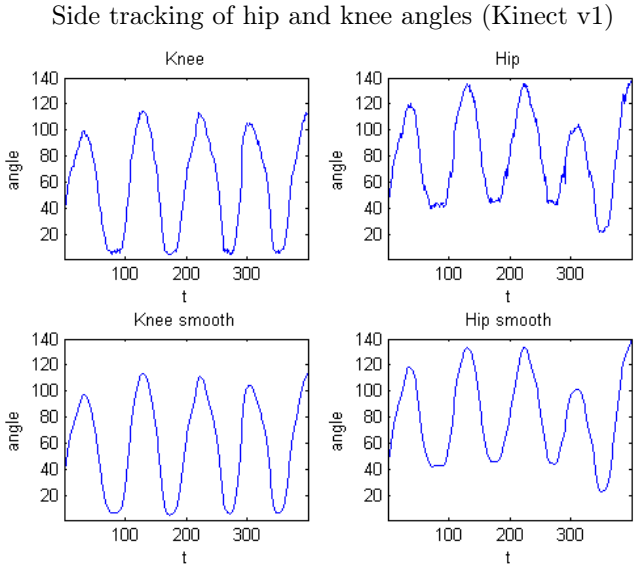

Figure 4.14: Knee and hip angle plot of the result of the side tracking algorithm. The bottom plots represent the knee and hip angle displacement after applying a smoothing filter.

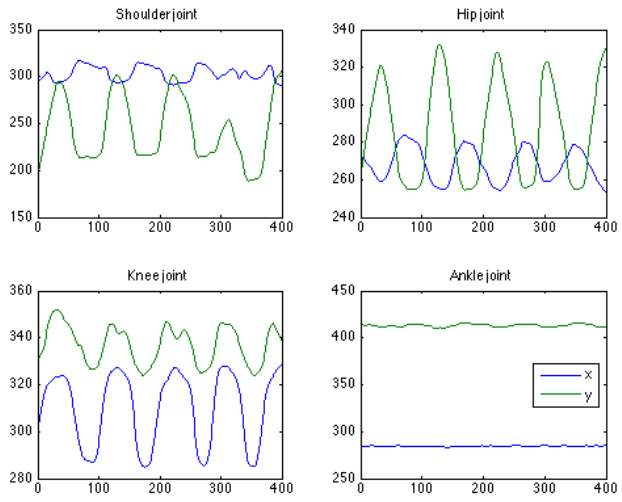

Figure 4.15: Plot of the x-and y-positions from the side tracking algorithm. The joints positions; shoulder (torso), hip, knee and ankle; are estimated using mean shift.

Table 4.6: *Session 1*: Statistical Analysis of markerless (MLS) and marker based (MBS) kinematics, during three tasks at two time points; vertical drop jump (VDJ), two-legged squat (squat) and single-leg squat (SLS); peak knee flexion (PF), frontal plane projection angle (FPPA) and knee-to-ankle separation ratio (KASR); at initial contact (IC) and peak flexion (PF). The values marked in red indicate a significant difference between the MLS and MBS for the compared means, and green means no significant difference was found (two-tailed paired samples t-test, $p < 0.05$). Spearman's Rank correlation coefficient (SRCC) values are marked green for results that show a significant correlation between variables (*/** = Correlation is significant at the 0.05/0.01 level, 2-tailed)

Statistical Analysis Session 1; MLS (Kinect v1) vs MBS (Qualisys)												
Comparing Means; Paired Samples T-test							SRCC		Intra-class Correlation Coefficient			
Variables			MLS vs MBS		MLS	MBS	MLS vs MBS		MLS (Kinect v1)		MBS (Kinect v2)	
Task	Kinematics	Time point	L	R	L - R	L - R	L	R	L	R	L	R
VDJ	PF		0.077	0.010	0.003	0.143	0.769*	0.599	0.376	0.519	0.685	0.626
	FPPA	at IC	0.186	0.313	0.250	0.274	0.657*	0.420	0.751	0.856	0.799	0.878
		at PF	0.069	0.293	0.001	0.162	0.783*	0.678*	0.695	0.587	0.624	0.599
	KASR	at IC	0.078				0.392		0.447		0.825	
		at PF	0.05				-0.252		0.153		0.662	
Squat	PF		0.00	0.001	0.064	0.279	0.600	0.783*	0.806	0.885	0.895	0.814
	FPPA	at PF	0.18	0.626	0.01	0.384	0.758*	0.794*	0.827	0.914	0.891	0.951
	KASR	at PF	0.016				0.927**		0.861		0.953	
SLS	PF		0.382	0			0.217	0.105	0.657	0.296	0.720	0.216
	FPPA	at PF	0.847	0.074			0.364	0.077	0.154	-0.066	0.640	0.467

Chapter 5

Discussion

Based on the implications of the results discussed in the previous chapter, we were able to answer the research questions stated in the introduction of this thesis. We set some requirements on the results that helped us determine if the MLS was accurate enough to analyze knee motion. The evaluation was done for each task;

- All variables had to show no significant difference from the MBS measurements.
- There had to be good correlation between the mean variables
- Good inter-trial reliability

Visual inspection of plotted graphs confirmed (or rejected) the notion that high correlation scores were the result of existing relationships between the MLS and MBS variables, and not just due to chance. The relatively small sample size has to be considered (especially for session 2), and influenced the confidence of the results.

5.1 Accuracy and Tracking Ability of Kinect MLS

Errors in the data from the two-legged squat task for two of the subjects caused us to exclude their trials in session 1 (Kinect v1). This meant that the statistical analysis for this task was performed using data from ten subjects instead of twelve. Another error in the two-legged squat data was found in the MBS (Qualisys) recording for a third subject. This however, affected only one of the trials so we used four values instead of five for all the ten remaining subjects. Despite the missing data, we decided that ten subjects and four trials for the two-legged squat would give an acceptable amount of data for the analysis. The results still provided enough information for us to draw a conclusion related to the objectives of the thesis.

5.1.1 Evaluation of Knee Motion - Vertical Drop Jump

The aim for this task was to evaluate if Kinect is accurate enough to evaluate knee motion during vertical drop jump (VDJ). The FPPA and knee flexion

angular displacement during vertical drop jump (VDJ) was plotted in figure 4.2. Since the data is not synchronized in time it is not an optimal comparison, but the plots show a good indication of similar results from MLS and MBS. The VDJ is definitely a good task for evaluating the FPPA and knee flexion. The statistical results for the VDJ in table 4.6 supports the impression we got from the plotted measurements. FPPA at initial contact (IC) show no significant difference from MBS, and has moderate-to-good rank correlation scores. A peak flexion (PF), FPPA was good, except for the significant difference between the left and right leg for MLS. We evaluated this as unimportant since we have good correlation scores (SRCC), and plots that indicate a real relationship between the MLS and MBS variables (FPPA and PF). Mean PF for the right leg shows a significant difference for VDJ which is interesting, but the figure 5.3 on the next page illustrates a possible error in the knee tracking for the right leg.

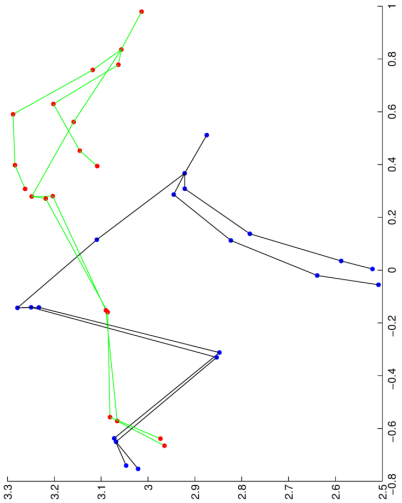

Figure 5.1: MLS skeletons; the green skeleton is at point of initial contact (IC) and the black is at peak flexion (PF) during a vertical drop jump. This subject has a high knee flexion angle at 100 degrees at PF.

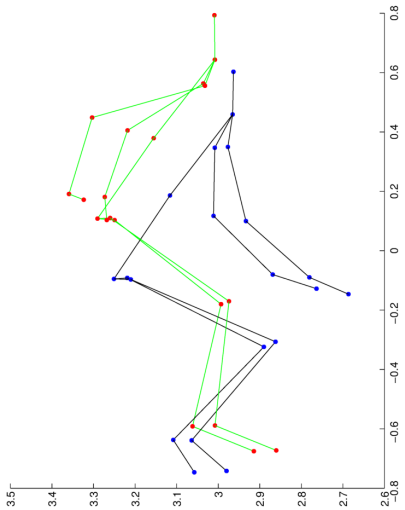

Figure 5.2: MLS skeletons; the green skeleton is at point of initial contact (IC) and the black is at peak flexion (PF) during a vertical drop jump. This subject has a moderate knee flexion angle of 90 degrees PF.

The illustrations in figure 5.1 and figure 5.2 show the skeletons for two different subjects during vertical drop jump in the sagittal plane. The green skeleton is plotted at point of initial contact (IC), and the black skeleton is at peak knee flexion (PF). The subject in figure 5.1 has a visibly larger PF angle for the right leg (100 degrees), than the subject in figure 5.2 (90 degrees). Frontal plane plots of the same skeletons are found in figure 5.3 on the next page and figure 5.4 on the following page. The right knee of the skeleton in figure 5.3 on the next page at PF is much more displaced than the left, and indicates a higher knee motion than the left leg. This could also be caused by

error in the tracking. Comparison with the skeleton in figure 5.4 supports the suggestion of higher knee motion and larger PF angle for the first subject. The skeletons indicate that the MLS can at least detect high knee knee motion based on the peak flexion angle at IC and PF for the VDJ. The statistical results from comparing MLS to the MBS are not strong enough in order to say that MLS is accurate. The moderate-to-good correlation scores for VDJ in session 1 for PF and FPPA ($0.42 - 0.78$ are promising, but a larger sample size is necessary to establish a real relationship.

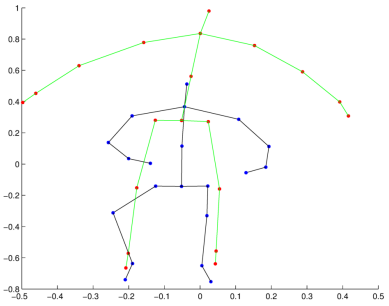

Figure 5.3: MLS skeletons; the green skeleton is at point of initial contact (IC) and the black is at peak flexion (PF) during a vertical drop jump. This subject has a high right knee flexion angle at 100 degrees at PF.

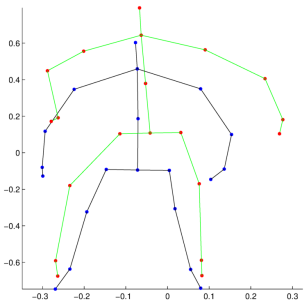

Figure 5.4: MLS skeletons; the green skeleton is at point of initial contact (IC) and the black is at peak flexion (PF) during a vertical drop jump. This subject has a moderate right knee flexion angle of 90 degrees PF.

5.1.2 Evaluation of Knee Motion - Front-oriented Squat

The aim for this task was to evaluate if Kinect is accurate enough to evaluate knee motion during front-oriented squat. The high correlation score (0.927) for the KASR during two-legged squat at PF, indicated a significant monotonic relationship between the mean values for the MLS (Kinect v1) and the MBS. The SRCC was also good for the squat task, but the PF and KASR showed a significant difference that implies inaccuracy compared to the MBS. It was not surprising that this specific variable gave a better result than the others, due to the fact that it does not rely on any hip joint center positions in the kinematic measurements. Therefore the problem with tracking these joints at deep positions due to self-occlusion, is eliminated. On the other side this variable does not give much insight to overall knee motion since the squat task does not include a lot of joint motion, and the KASR is expected to remain stable. FPPA is not a good measure of knee motion during squats since the thigh (femur) is reduced to a single point when the legs are parallel at peak flexion.

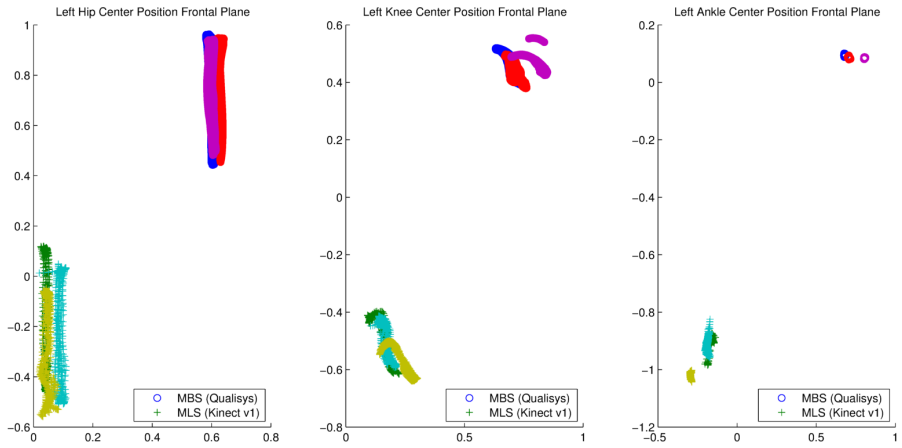

Figure 5.5: *Session 1*: Scatter plot of left leg joint centers in the frontal plane for three subjects; hip, knee and ankle during two-legged squat (front-oriented). The coordinate systems are inverse of each other for plot readability.

5.1.3 Evaluation of Knee Motion - Single-leg Squat

The aim for this task was to evaluate if Kinect is accurate enough to evaluate knee motion during single-leg squat. Spearman's rank correlation coefficient (SRCC) that evaluates the ranking relationship between the mean values, was very low and not significant for both legs in the SLS for all tasks. The mean values for the measurements from the Qualisys MBS recording of SLS were interestingly high for the right leg. This is seen in table 4.1 on page 72 where the mean and standard deviation are much higher, and we believe this was due to error since Qualisys is meant to be reliable in its measurements. The interpretation of the results from the statistical comparison for SLS can therefore not be trusted, and the poor correlation can in the best case be caused by an error from the MBS recording or tracking procedure.

Peak Flexion (PF)

The results showed a significant difference between the mean peak knee flexion (PF) for the SLS with the right leg, but not for the left leg. From an observational point of view the subjects seemed weaker in their right leg, and more than once the other leg caused a temporary occlusion that caused a temporary failure in the skeletal tracking of the single leg. During the single-leg squat (SLS) the subject had to keep the other leg above the floor, and not in front of the leg doing the squat. The technique varied noticeably from subject to subject, and the ones that struggled with keeping their balance in the SLS, leaned forward more in order to squat deep enough. The forward-leaning occludes more of the body, especially the hip area, and naturally these joint positions become harder to estimate.

Frontal Plane Projection Angle (FPPA)

In the results from the frontal plane projection angle (FPPA) for the SLS, we saw an indication of a better accuracy in the tracking of the left leg. The p-value for the left leg that was larger than the right, but neither indicated a significant difference. In the FPPA the variables showed no significant difference between the MLS and MBS systems. The FPPA measure is calculated using only the joint positions in the frontal plane, and the depth values are ignored. The paired samples t-test for FPPA gave a much better result for both legs in the SLS than for the PF measure, which indicates that the source of error in the MLS for the PF, is the third coordinate, depth. As for the two-legged the FPPA might not be the best measure for knee motion since the thigh is almost invisible in the frontal plane for the deepest position. Recording the task with the MLS at an angle that includes more of the side of leg performing the squat, could potentially increase the accuracy of the measurement.

5.1.4 Evaluation of Motion Tracking - Side-oriented Squat

The results from the new method for side-tracking using image analysis techniques, were not evaluated using statistical measurements. The plots of the estimated joint positions and angular displacement are discussed in terms of knowledge about the nature of the squat exercise.

Sources of Error

It was expected that the ankle position appeared as a straight line in the plot in figure 4.15 on page 86, since this is the only joint that should not move much during the squat. The little movement that was detected, could be due to the player putting more weight in the toes when she goes deeper into the position, causing the heel to slightly lift off the floor. The error in the knee joint tracking could probably be fixed with more filtering or adjustment of the algorithm. The edge of the "corner" of the knee has high gradients in two significantly different directions, and should be a robust feature for tracking [41]. The plots of the joint positions in figure 4.15 on page 86 show that the error appears when the player is on her way up from the squat. This explains that the key point in the knee changed appearance, and there was no longer a "corner" to track. The torso tracking suffered from the fact that the weight-plate was occluding the upper part of the body during parts of the video sequence.

Possible Improvements

A different matching strategy could improve the performance of the knee-tracking to account for the change in appearance of the key point. The size of the window around each joint and the number of iterations, were parameters that we experimented with to optimize the motion tracking of the joints. The final parameter value for the similarity threshold was 0.2 (0.16 initially), and the maximum number of iterations were lowered from 5 to 3. Since the same parameter values were used for all of the joints, the algorithm could be further improved by optimizing the parameters for each individual key point in the tracking. Some of the joint positions can handle more constraint than for example the area around the knee joint, which has a bigger change in appearance in the different

parts of the video sequence. Another problem with the knee area was that it is very homogeneous in color, so there appeared to be too many false positives in the tracking algorithm. This could be fixed with a lower threshold for this particular key point and possibly also a smaller window size. The performance of the algorithm could be evaluated by comparing the estimated angles to the angles obtained from the marker-based motion capture system (Qualisys).

Limitations

A flaw to the tracking algorithm is that it only considers the joint positions in the image (x,y) plane and not the real world coordinates in 3D. Therefore there could be an error in the calculations, since the side of the body of the player doing the squat task is not exactly in a plane. The feet and legs are usually placed wider than the width of the hips and shoulders.

The proper technique for a two-legged squat is individual and depends not only on the physical abilities and features of each person, but also on the purpose of the training. Athletes from different sport does not have the exact same need for muscle strength. In basketball for example, most of the time the players jump from a close-to-standing position and need to quickly elevate without having the time to bend into a deeper squatting position before leaving the ground.

5.1.5 Inter-trial Reliability for Kinect v1

The reliability was evaluated between trials using the results from the intra-class correlation coefficient (ICC). The two-legged squat showed the best overall inter-trial reliability according to the ICC values for all variables in session 1. The Kinect v1 had a reliable tracking of joint positions between trials for all kinematic variables in slow movements like the two-legged squat, based on the correlation coefficients in table 4.6 on page 87. The ICC for the VDJ and SLS varied between poor and good correlation, the squat showed very good correlation overall. This could mean that the Kinect v1 sensor deals better with predictable movements without much variation (like the two-legged squat) in terms of accurate joint tracking.

5.1.6 Evaluation of Accuracy of Kinect v2

In session 2 with Kinect v2 as the MLS, we had data from the vertical drop jump (VDJ) and the two-legged squat (front-oriented) in the statistical analysis. The results are discussed separate from session 1 since the data sample size was too small to give anything other than an impression of the accuracy of Kinect v2. One of the trials in the VDJ task had to be dropped, due to error in the tracking for one of the subjects (MBS).

The knee flexion angles during a SLS (left leg) from session 2, is plotted to the right in figure 5.6 on the following page. Due to the artifacts in the red curve (Kinect v2) during the time of peak flexion, it was not possible to extract the peaks using the same method from session 1. There is a clear difference when comparing figure 5.6 on the next page to the plots in figure 4.10 on page 82 from session 1.

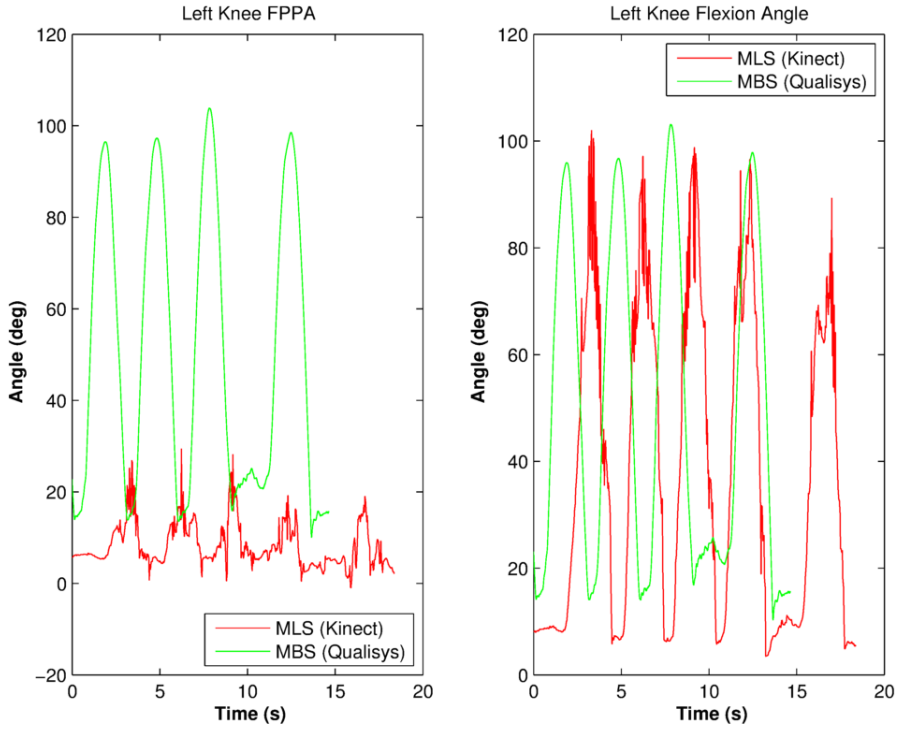

Figure 5.6: *Session 2*: Comparison of MLS (Kinect v2) and MBS Knee flexion and frontal plane projection angles (FPPA) during five consecutive single-leg squat (SLS) trials. The MLS show severe artifacts in the plotted curve at the times of peak knee flexion. The FPPA measure have peak values significantly smaller than the MBS measure. (Not synchronized in time.

The results from the statistical analysis for session 2 were presented in table 4.3 on page 75. It is important to keep in mind that the small number of subjects in this pilot-sessions makes it hard to state any conclusion on the performance of the Kinect v2 sensor based on the statistical results.

Spearman's correlation coefficient (SRCC) gave consistent p-values of either 1.0, 0.5, or -0.5 for the vertical drop jump variables seen in table 4.2. If the sample size had been larger, these correlation values would be interpreted as moderate to excellent.

Kinect v2 - Vertical Drop Jump

The scatter plot in figure 5.8 from session 2 for mean peak flexion, illustrates the fact that with only three points the SRCC value of 1.0, is easily attributed to chance instead of a strong correlation between the MLS and MBS. In order to evaluate if there actually is a relationship between the variables, we take a closer look at the plots in figure 4.4 of the peak flexion during VDJ for one of the subjects. The bottom plots representing the knee flexion for each leg are shaped very similar and in this case, the peak flexion (PF) angle is larger for the MLS (Kinect v2) than for MBS. The mean values in table 4.2 on page 73

for PF during VDJ confirms that this is the trend for all subjects. The opposite is the case for the FPPA, where the mean angle is extremely lower for the MLS than for the MBS. This is not surprising when observing the FPPA plot in figure 4.4 on page 78, which indicates that we failed to obtain proper results for this measure.

The scatter plots of the lower body joint centers in figure 5.7, can contribute to an explanation of why the FPPA measure failed. The positions are plotted for the right leg during the contact phase of VDJ, MLS suffer from a lower frequency with a substantial number of fewer frames per second. It looks like the right knee joint (MLS) is not accurately tracked in the lateral direction in the frontal plane although there is a small tendency of the "collapse" of the knee at the lowest position. This skew is easier to recognize in the plotted knee position for MBS. In the theory chapter the FPPA was presented with an illustration that explained how it measures the projected knee angle in the frontal plane. The lateral motion of the knee joint in the frontal plane was not tracked to the extent needed for a successful measure of the FPPA, and the result is the curve for Kinect v2 MLS seen in figure 4.4 on page 78.

Session 2- lower body joint center positions (right leg) during VDJ

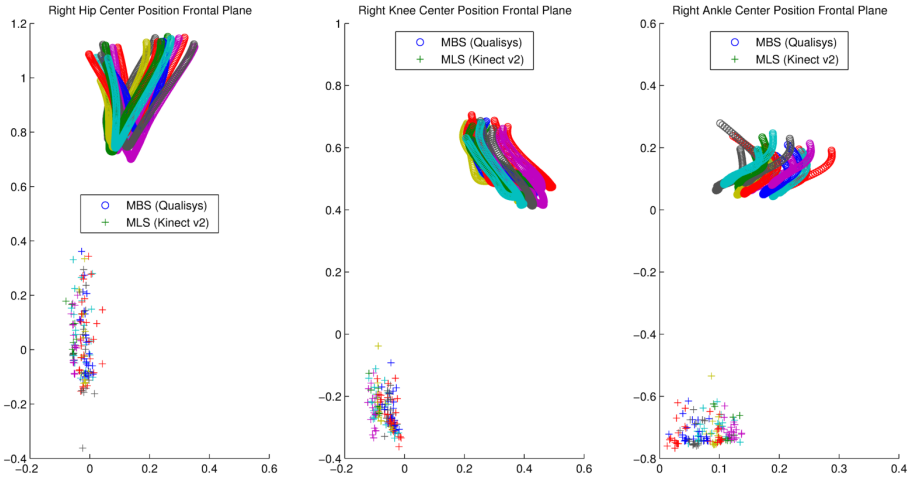

Figure 5.7: *Session 2*: Scatter plot of right leg joint centers; hip, knee and ankle during vertical drop jump. MBS values at the top and Kinect v2 MLS in the bottom left. There is a larger displacement in the lateral directions in the MBS joint centers than in the The MBS knee joint "collapses" at peak flexion (PF) indicated by the skew to the right, but the same is not seen to the same extent in the Kinect v2 positions. The coordinate systems are inverse of each other for plot readability.

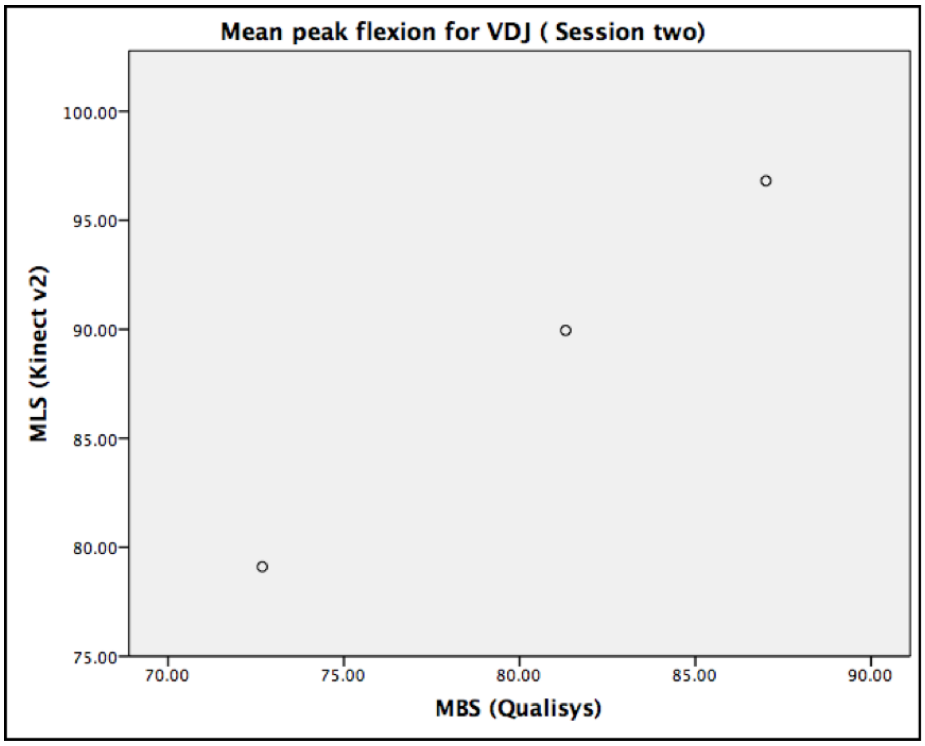

Figure 5.8: *Session 2*: The scatter plot from session 2 for mean peak flexion for vertical drop jump. Spearman's correlation coefficient was 1.0 for this variable, but does not prove a strong correlation between MLS (Kinect v2) and MBS (Qualisys) due to small sample size.

Kinect v2 - Front-oriented squat

The descriptive statistics in table 4.2 on page 73 for Kinect v2 MLS were very promising with mean values for peak flexion (PF) almost equal to Qualisys MBS. In the statistical comparison we found no significant difference between these variables for both legs during the front-oriented squat, seen in table 4.3 on page 75.

FPPA at PF was on the other hand significantly different, and as mentioned earlier it is not a good measure for the squat since the thigh becomes a single point in the frontal plane at the deepest position. The positions for the three lower body joints; left hip, left knee and left ankle during the two-legged squat are plotted in the frontal plane in figure 5.9 for session 2. The vertical variation in the knee and ankle position can be explained by the difficulty Kinect v2 have of detecting the body parts (thigh and shank), at the deepest position in the frontal plane. The depth position is apparently not affected by the occlusion to the same degree, since the 3D knee flexion for Kinect v2 is very accurate and not significantly different from MBS. Knee flexion angle (left and right knee) for one subject during the front-oriented squat, was plotted in figure 4.8, with seemingly very good correlation between MLS and MBS.

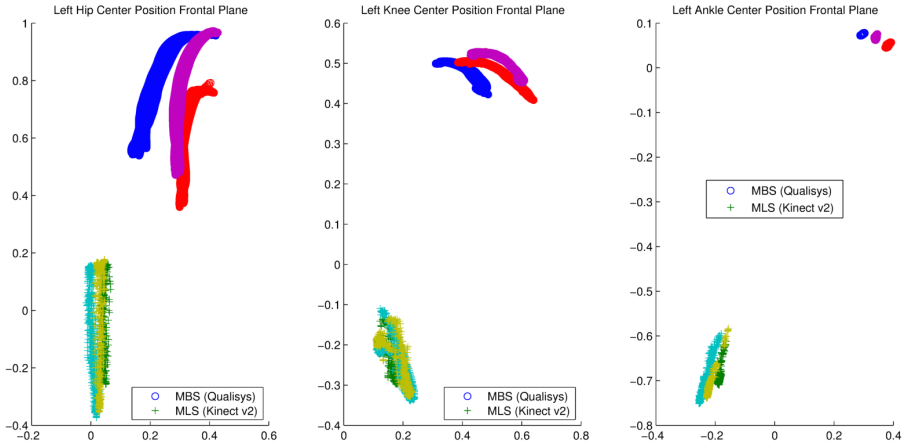

Figure 5.9: *Session 2*: Scatter plot of left leg joint centers in the frontal plane for three subjects; hip, knee and ankle during two-legged squat (front-oriented). The coordinate systems are inverse of each other for plot readability.

Kinect v2 - Single-leg squat

In session 2 we saw more noise in the kinematic measurements of knee flexion and frontal plane projection angle (FPPA) for the single-leg squat, than in session 1. The knee flexion measure uses 3D coordinates in its calculation and could be influenced by error in the depth estimation. For the three subjects in session 2 we compared the joint center positions in the frontal plane, during the single-leg squat (left leg). The scatter plots for the hip, knee and ankle joint centers during SLS (left leg), are shown in figure 5.11 on the following page for session 2, and in figure 5.10 for session 1. In the ankle joint center positions we see a larger vertical variability in the MLS (Kinect v1 and v2) than the MBS. The knee flexion angle, seen in the sagittal plane in figure 3.10 on page 56 and the FPPA measure illustrated in figure 2.20 on page 39, are both estimated based on the position of the hip, knee and ankle joints during the SLS. The variation in ankle position will affect the calculations of knee flexion angle and the FPPA, decreasing the accuracy.

There is a significant difference in the variance in the ankle joint center position between the MLS and MBS, while the hip and knee have very similar displacement. The Kinect v2 MLS has a even larger error in the vertical position than Kinect v1, compared to Qualisys MBS in each task. From visual inspection of the plots, the ankle position estimated by Kinect v2 MLS, differ with approximately 15 cm from the starting position to the point of peak flexion. The Kinect v1 has a corresponding estimated average difference of about 10 cm. This could explain the lack of smoothness in the curve for the Kinect v2 knee flexion and FPPA. Kinect v2 has a higher depth camera resolution (512x424), which can contribute to a higher sensitivity when the raw data is used in the calculations of the kinematic measurements. The skeleton tracking in Kinect v2 is expected to take advantage of better depth estimation with the time-of-flight technology, but looks to still have problems with tracking the ankle joint when

the shank body segment is not visible in deep squat positions. In general self-occluding is an issue for both sensors. The testing protocol for the screening sessions defined the MLS sensor to be placed directly in front of the subject. This position is good for real-time observation of knee control in vertical drop jump, but we expect that placing the Kinect sensor at an angle (30 degrees) during recording of squatting tasks, will improve the skeletal tracking by reducing the risk of self-occluding of body parts. The FPPA calculation might not benefit from the change in angle for the recording. Another issue with the screening setting is that the Kinect v2 sensor had to be placed further away than necessary, in order not interfere with the MBS Qualisys camera acquisition.

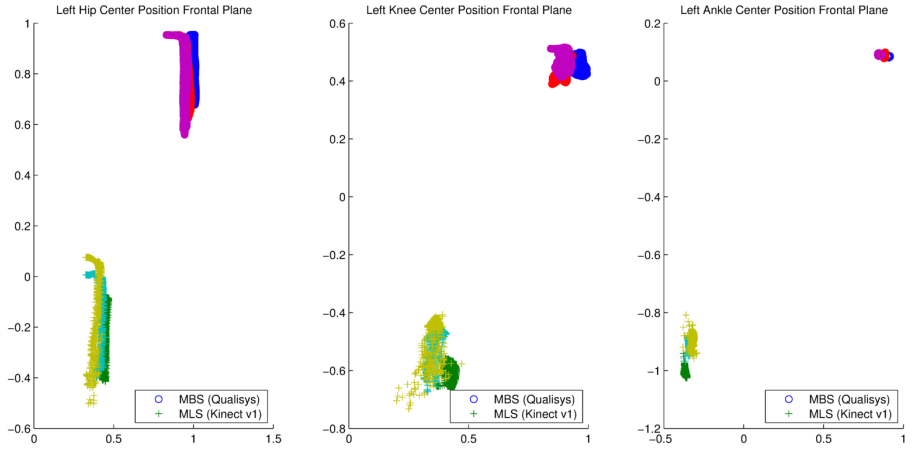

Figure 5.10: *Session 1*: Scatter plot of left leg joint centers in the frontal plane for three subjects; hip, knee and ankle during single-leg squat. The coordinate systems are inverse of each other for plot readability.

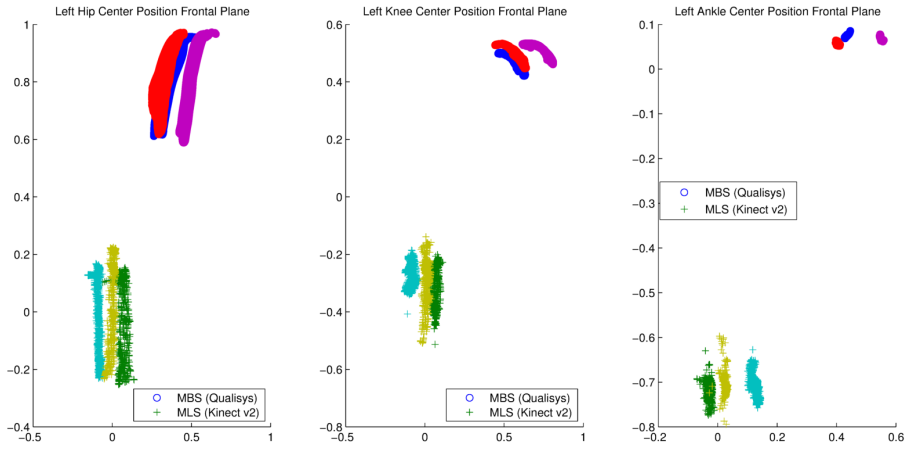

Figure 5.11: *Session 2*: Scatter plot of left leg joint centers in the frontal plane for three subjects; hip, knee and ankle during single-leg squat.The coordinate systems are inverse of each other for plot readability.

Side Tracking with Kinect v2

According to the Offset Joint Regression method described in 2.2.2 on page 29, the Kinect v2 sensor was believed to track even occluded body parts. During the side-squat task for the first subject in the second session (Kinect v2), the joints and skeleton for the lower body were tracked for the side facing the sensor. At the time of recording with Kinect Studio 2.0 this seemed very promising and a screenshot of the playback of this particular case can be seen in figure 5.12. When we tried to save the body joint data from this recording, the application failed to track the skeleton and no raw data could be obtained.

Figure 5.12: Kinect Studio recording of the two-legged squat from the side. The joints and skeleton on the side towards the sensor are tracked (thick lines), but the internal joints that are occluded are inferred (thin lines).

The method developed in this thesis could not be run on the Kinect v2 images from the Kinect Studio recordings, as the RGB color frames were not yet available for streaming through the application.

Kinect v2 - Inter-trial Reliability

The intra-class correlation coefficients (ICC) for the MLS (Kinect v2) and MBS (Qualisys) during four trials of VDJ were presented in table 4.4. The small sample size made it difficult to evaluate the inter-trial reliability based on these results. It is expected that MBS has good correlation between trials, so the high ICC value can reflect a real significant relationship. The MLS has a ICC value of 0.472 for the left leg during the VDJ. Without any knowledge about the general reliability of Kinect v2, it is not enough to claim a good correlation between the trials. Since this statistical measure is based on few observations, the only hope is that an even better correlation is obtained with a bigger sample size. More testing with the Kinect v2 is needed to see if the ICC values for the VDJ in table 4.4 on page 75 and table 4.5 on page 75 are representative of the inter-trial reliability of the Kinect v2. The front-oriented squats had very good correlation between trials, and looking at the plotted curves, the Kinect v2 seem to be reliable in its measurements for all subjects.

5.2 Synchronization in Time and Space

The objective that we failed to solve, was the synchronization in time and space for the two systems with a large difference in frame rate. A patent by Berge from 2011 [8] presents a way to synchronize external events to video stream by using a triggered event to activate a light source that is attached to the camera. The method can be used with the Kinect sensor's camera which is a monochrome CMOS sensor that reads out the image line by line. A LED-light attached right next to the CMOS sensor on the Kinect could at the time of activation completely cover the field of view of the Kinect with a blinding light. The time of the activation of the LED-light can be detected at a higher frame rate than the Kinect sensor's original frame rate. This is done by utilizing the interaction between the line readout from the CMOS-sensor, and the continuous integration time in the CMOS chip. The difference between consecutive line readouts can be compared to find the exact frame for when the light was activated, since the lines after the activation of the light will increase in brightness. This gives a signal that can be used for synchronization with a possible precision of:

$$\frac{1}{(framerate * horizontallines)persecond} = \frac{1}{(30 * lines)second}$$

This method will give an accurate time synchronization, but it is also not the easiest to implement. An optimal method would code the time synchronization into the video stream of the Kinect without having to physically manipulate the sensor.

Alignment of 3D Models A few different methods for finding the transformation between the MBS and the MLS coordinate systems were experimented with, but ended up not being accurate enough when implemented on the 3D data. With the Kinect v2 testing, the orientation data inspired the use of quaternions to find the rotation between the body pose of the two systems we wanted to compare. More time to experience with this will probably contribute to finding a robust method of aligning the skeletons in space.

Chapter 6

Conclusion

The results from the evaluation of Kinect MLS as a tool for analyzing sports movement, can be summed based on the objectives for this thesis:

- Evaluation of the accuracy and tracking ability of MLS, Kinect v1 and v2
  - Kinect v1 and v2 accurately estimated peak flexion during the two-legged squat task based on high correlation scores and visual inspection.
  - Plots of PF and FPPA for VDJ and the front-oriented squat, showed that the general shape of the curves of most of the squat-variables, resembled those from the MBS, and confirmed the indicated relationship between Kinect v1 MLS and MBS variables.
  - KASR was very varying in its results for different tasks, but was estimated accurately for VDJ and squat by Kinect v2
  - Accuracy was poor for the single-leg squat (session 1), and the only measure with no significant difference was FPPA at PF for the left leg.
  - SLS scores for MBS were not as expected for the right leg, and indicated error in the processing phase. The task was not included in the results from session 2 due to inability to extract correct peak flexion values..
  - Adjustments to screening protocol will most likely improve the accuracy of the skeletal tracking (front and side) for Kinect MLS. Kinect v2 show improved ability of depth estimation and skeletal tracking in 3D, compared to Kinect v1.
- Investigation of the inter-trial reliability of MLS for each task
  - The two-legged squat showed the best overall inter-trial reliability according to good and very good intra-class correlation for all variables in session 1 and session 2.
  - For the VDJ, inter-trial reliability was best for the FPPA measure at IC and PF (moderate and good correlation)

- Assessment of methods for synchronizing the data from the MBS and MLS in time and space
  - The 3D data from MLS and MBS was difficult to synchronize in time due to large difference in frame rate. A simple method for estimating a common start-point in time for MLS and MBS, was found for session 1.
  - Synchronization in 3D space for kinematic measures, require defining local coordinate systems for each body segment in MLS skeleton. Orientation represented by quaternions for MLS joints, can be used to find the rotation between the coordinate systems
- Investigation of the side tracking ability of MLS
  - Inconclusive results since no skeleton data from side tracking was compared to the MBS data.
  - Side tracking with Kinect v2 looked promising, but we failed to acquire skeleton data with Kinect v1 and v2 for the side-oriented squat. Adjustments have to be made in order to track the joints properly for comparison with MBS.
- Considerations for use of MLS in a clinical setting for knee motion assessment
  - Further testing and verification of the accuracy of the Kinect MLS sensor has to be done before it can be implemented in a clinical setting.

Kinect v1 seemed to underestimate the peak flexion (PF) in all trials for the two-legged squat. This can indicate systematic error that in the best case can be corrected. Results from session 2 showed improvement in PF measurements with accurate values compared to MBS. Kinect v2 seems to be a very good upgrade from the first version of the sensor. From an observational point of view, the Kinect v2 also seemed much more stable in its skeletal tracking throughout the second screening session. Future experiments are necessary for validating the accuracy of Kinect v2 MLS for use in ACL injury risk assessment.

6.1 Further Work

The study done in this thesis used a very small sample size of the target population, and further work would be to validate the Kinect v2 sensor using a larger sample size and a better synchronization method. The orientation of the coordinate system in the MBS can be defined to match the Kinect skeleton for easier calculation of the transformation between the position of the 3D models. Changes in the testing protocol and physical placement of the MLS can improve the quality of the tracking ability and accuracy. Kinect v2 should be placed closer to the subject to take advantage of the increased field of view, and for better depth estimation and skeletal tracking.

It would be of interest to develop a joint coordinate system (JCS) for the lower body bone segments like the one implemented for the Qualisys MBS. This requires successful acquisition of skeletal joint positions in the static trial for the MLS. Successful recordings of the static trial for all subject, would have increased the motivation for completing a method for transforming the Kinect data for alignment with the MBS in space. With a proper JCS definition for Kinect data, a more realistic comparison can be done with definitions of angular displacement that better describe human motion than our 3D measures.

6.1.1 Improved Depth Estimation

Comparisons we did of scatter plots of the hip, knee and ankle joint centers in the frontal plane, indicated that a cause of error in the calculated measurements was inaccurate tracking of the hip joints in the for Kinect v1. Insecurity in the hip joint center locations could be due to depth perception noise. In a study by Raptis et al. [35] the Kinect v1 hip tracking ability was affected by noise in depth perception. The depth coordinate of joints seemed to be tracked accurately by the Kinect v2, compared to coordinates only for the frontal plane. We base this conclusion on the small difference in mean values for the peak flexion during the two-legged squat, and the poor result obtained for the projection angle in the frontal plane.

Recording the squatting tasks at a an angle, for example between 20 and 40 degrees to the side of the subject, could improve the tracking of the knee and ankle joints. If the skeleton is successfully tracked without occlusion, the adjusted recording angle could also improve the side tracking ability of the sensor. Extracting all joint positions need for calculating the back, hip and knee flexion angle for both legs in one recording, would be beneficial for screening purposes. This would also be applicable for the single-leg squat to avoid self-occlusion. A smoothing filter applied to the MLS data before extracting the kinematic measurements, could solve the problem of not being able to extract the peaks in the data from the single-leg squatting task with Kinect v2 in the second session.

Comparing the hip angle from MLS to the MBS, requires the center position of the torso. Defined the dynamic position of the torso in the kinematic output from the marker based system for all trial, would make it possible to evaluate the hip angle from the MLS 3D data. The hip angle is a good indicator of squatting technique, and comparing the results from our side-tracking algorithm developed in this study to MBS, would also be an interesting next step.

6.1.2 Clinical Setting

The Kinect MLS would be easy to implement in a clinical setting in a practical point of view. The sensor is small in size, and the Kinect v2 has a larger field of view and requires less distance to the user. In order to successfully use MLS in a clinical setting, a comparison could be done between the MLS and a 2D video system. The KASR is a measure that can be defined in the frontal plane in 2D video sequences. Physiotherapist have been proven as able to identify female athletes with high valgus knee angles, Nilstad et al. [32]. If the Kinect is validated by comparing it to real-time observational screening, good results will support the promising result from our study. A good correlation with observational screening also implies positive considerations for using Kinect MLS in a clinical setting.

Bibliography

[1] Eva Ageberg et al. "Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat". In: *BMC musculoskeletal disorders* 11.1 (2010), p. 265.

[2] B Bonnechere et al. "Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry". In: *Gait and Posture* 39.1 (2014).

[3] Fritz Albregtsen. *INF4300 - Digital Image Analysis: Hough Transform*. 2013. URL: http://www.uio.no/studier/emner/matnat/ifi/INF4300/h13/undervisningsmateriale/inf4300-2013-f03-hough.pdf.

[4] Douglas G Altman. "Some common problems in medical research". In: *Practical statistics for medical research* 1 (1991), pp. 396–403.

[5] M Azimi. "Skeletal Joint Smoothing White Paper". In: *MSDN digital library* (2012).

[6] Roger Bartlett. *Introduction to sports biomechanics*. Taylor & Francis, 1997.

[7] AL Bell, DR Pedersen, and RA Brand. "A comparison of the accuracy of several hip center location prediction methods." In: *Journal of biomechanics* 23.6 (1990), p. 617.

[8] A. BERGE, J.T. Thielemann, and K. Haugholt. "Trigger device". WO Patent App. PCT/EP2010/057,324. Dec. 2011. URL: http://www.google.com/patents/WO2011147454A1?cl=en.

[9] Asbjørn Berge. *INF5300 Advanced Topics in Digital Image Analysis: Statistical Tracking*. 2013. URL: http://www.uio.no/studier/emner/matnat/ifi/INF5300/v13/undervisningsmateriale/inf5300_v2013_lecture9_tracking_6pp.pdf.

[10] Sylvain Bernhardt. *Mean-Shift Video Tracking*. 2008. URL: http://www.mathworks.com/matlabcentral/fileexchange/35520-mean-shift-video-tracking.

[11] Peter Bone. *Hough Transform for circle detection*. 2006. URL: http://www.mathworks.com/matlabcentral/fileexchange/9833-hough-transform-for-circle-detection/content/houghcircle.m.

[12] Oslo Sports Trauma Research Center. *Risk factors for ACL-injuries in elite female football players*. 2009-2014. URL: http://www.ostrc.no/en/Project/160---Risk-factors-for-ACL-injuries-among-elite-female-football-players/.

[13] Roy B Davis III et al. “A gait analysis data collection and reduction technique”. In: *Human Movement Science* 10.5 (1991), pp. 575–587.

[14] Elham Dolatabadi et al. “A MARKERLESS MOTION TRACKING APPROACH TO UNDERSTAND CHANGES IN GAIT AND BALANCE: A CASE STUDY”. In: ().

[15] Janice J Eng and David A Winter. “Kinetic analysis of the lower limbs during walking: what information can be gained from a three-dimensional model?” In: *Journal of biomechanics* 28.6 (1995), pp. 753–758.

[16] B. Freedman, Y. Arieli, and A. Shpunt. *Distance-varying illumination and imaging techniques for depth mapping*. WO Patent App. PCT/IL2008/000,838. Feb. 2010. URL: https://www.google.com/patents/WO2008155770A3?cl=en.

[17] B. Freedman et al. *Depth mapping using projected patterns*. WO Patent App. PCT/IL2008/000,458. Feb. 2010. URL: https://www.google.com/patents/WO2008120217A3?cl=en.

[18] Edward S Grood and Wilfredo J Suntay. “A joint coordinate system for the clinical description of three-dimensional motions: application to the knee”. In: *Journal of biomechanical engineering* 105.2 (1983), pp. 136–144.

[19] C Herrera, Juho Kannala, Janne Heikkilä, et al. “Joint depth and color camera calibration with distortion correction”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34.10 (2012), pp. 2058–2064.

[20] Timothy E Hewett et al. “Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes A prospective study”. In: *The American journal of sports medicine* 33.4 (2005), pp. 492–501.

[21] Hideyuki Koga et al. “Mechanisms for noncontact anterior cruciate ligament injuries knee joint kinematics in 10 injury situations from female team handball and basketball”. In: *The American journal of sports medicine* 38.11 (2010), pp. 2218–2225.

[22] Eirik Kristianslund et al. “Sidestep cutting technique and knee abduction loading: implications for ACL prevention exercises”. In: *British journal of sports medicine* (2012).

[23] Daniel Lau. *The Science Behind Kinects or Kinect 1.0 versus 2.0*. 2013. URL: http://www.gamasutra.com/blogs/DanielLau/20131127/205820/The_Science_Behind_Kinects_or_Kinect_10_versus_20.php.

[24] Jennifer L McGinley et al. “The reliability of three-dimensional kinematic gait measurements: a systematic review”. In: *Gait & Posture* 29.3 (2009), pp. 360–369.

[25] SG McLean et al. “Evaluation of a two dimensional analysis method as a screening and evaluation tool for anterior cruciate ligament injury”. In: *British journal of sports medicine* 39.6 (2005), pp. 355–362.

[26] Ryan L Mizner et al. “Comparison of Two-dimensional Measurement Techniques for Predicting Knee Angle and Moment during a Drop Vertical Jump”. In: *Clinical Journal of Sport Medicine* 22.3 (2012), p. 221.

[27] Thomas B Moeslund et al. *Visual Analysis of Humans*. Springer, 2011.

[28] Gregory D Myer, Kevin R Ford, and Timothy E Hewett. "New method to identify athletes at high risk of ACL injury using clinic-based measurements and freeware computer analysis". In: *British Journal of Sports Medicine* 45.4 (2011), pp. 238–244.

[29] Gregory D Myer et al. "Biomechanics laboratory-based prediction algorithm to identify female athletes with high knee loads that increase risk of ACL injury". In: *British journal of sports medicine* (2010), bjsports69351.

[30] Gregory D Myer et al. "Development and validation of a clinic-based prediction tool to identify female athletes at high risk for anterior cruciate ligament injury". In: *The American journal of sports medicine* 38.10 (2010), pp. 2025–2033.

[31] Microsoft Developer Network. *Natural User Interface - Skeletal Tracking*. 2013. URL: http://msdn.microsoft.com/en-us/library/hh973074.aspx.

[32] Agnethe Nilstad et al. "Physiotherapists Can Identify Female Football Players With High Knee Valgus Angles During Vertical Drop Jumps Using Real-Time Observational Screening". In: *journal of orthopaedic & sports physical therapy* 44.5 (2014), pp. 358–365.

[33] Odd-Egil Olsen et al. "Injury mechanisms for anterior cruciate ligament injuries in team handball a systematic video analysis". In: *The American Journal of Sports Medicine* 32.4 (2004), pp. 1002–1012.

[34] Alexandra Pfister et al. "Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis". In: *Journal of medical engineering & technology* (2014), pp. 1–7.

[35] Michalis Raptis, Darko Kirovski, and Hugues Hoppe. "Real-time classification of dance gestures from skeleton animation". In: *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM. 2011, pp. 147–156.

[36] Jamie Shotton et al. "Efficient human pose estimation from single depth images". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.12 (2013), pp. 2821–2840.

[37] Jamie Shotton et al. "Real time human pose recognition in parts from single depth images". In: *Communications of the ACM* 56.1 (2013), pp. 116–24.

[38] Inge Söderkvist and Per-Åke Wedin. "Determining the movements of the skeleton using well-configured markers". In: *Journal of biomechanics* 26.12 (1993), pp. 1473–1477.

[39] Silje Stensrud et al. "Correlation between two-dimensional video analysis and subjective assessment in evaluating knee control among elite female team handball players". In: *British Journal of Sports Medicine* 45.7 (2011), pp. 589–595.

[40] Erik E Stone et al. "Evaluation of the Microsoft Kinect for screening ACL injury". In: *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE. 2013, pp. 4152–4155.

[41] Richard Szeliski. *Computer vision: algorithms and applications*. Springer, 2010.

[42] Eric Vittinghoff, Stephen Shiboski, Charles E McCulloch, et al. *Regression methods in biostatistics*. Springer, 2005.

[43] Herman J Woltring. "Smoothing and differentiation techniques applied to 3-D data". In: *Three-dimensional analysis of human movement* (1995), pp. 79–99.

[44] Zhengyou Zhang. "Microsoft kinect sensor and its effect". In: *Multimedia, IEEE* 19.2 (2012), pp. 4–10.